Characterization of Two G-Protein Coupled Receptors and One Fox Transcription Factor in Drosophila Embryonic Development

By
Caitlin D. Hanlon

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
July 2015

ABSTRACT

Cell migration is an exquisitely intricate process common to many higher organisms. Variations in the signals driving cell movement, the distance cells travel, and whether cells migrate as individuals, clusters, or as intact epithelia are all possible. Cell migration can be beneficial, as in development or wound healing, or detrimental, as in cancer metastasis. To begin to unravel the complexities inherent to cell migration, the Andrew lab uses the Drosophila salivary gland as a relatively simple model system for learning the molecular/cellular events underlying cell movement. The salivary gland begins as a placode of polarized columnar epithelial cells on the surface of the embryo that invaginates and move dorsally until a turning point is reached. There, it reorients and begins posterior migration, which continues until the gland reaches its final position along the anterior-posterior axis of the embryo. The broad goal of my work is to identify and characterize other key players in salivary gland migration. I characterized two G-protein coupled receptors (GPCRs) – *Tre1* and *mthl5* – which are expressed dynamically in the embryo. By creating a null allele of *Tre1*, I found that *Tre1* plays a key role in germ cell migration and affects microtubule organization in the migrating salivary gland. I created a *mthl5* mutant allele using the CRISPR/Cas9 system. *mthl5* plays a role in the cell shape changes that drive salivary gland invagination. I have identified a potential ligand of Mthl5, *fog*, which plays a known role in mediating cell shape changes. Mutant alleles of *fog* phenocopy mutant alleles of *mthl5*. In a separate project, I characterized the role of the Fox transcription factor *FoxL1* in the *Drosophila* embryo. Mis-expression of *FoxL1* causes severe defects in salivary gland migration and muscle organization. I found that *FoxL1* is upstream of the signaling molecule *sema2a* and plan to identify more targets through microarray analysis. Together, these data provide important information for how tissues integrate signaling information to arrive at the correct final destination.

ACKNOWLEDGEMENTS

A special thanks to all of those who have supported me throughout my journey at Hopkins. Without the people mentioned below, none of my accomplishments would have occurred, and life certainly would not be as fun.

Thanks to my parents for coming to visit and bringing the rest family – "therapy" dogs included – to visit regularly. While my parents would check in regarding my work, I especially appreciated being able to just sit quietly with them in front a campfire. Thanks to my sister for listening to all of my science complaints and miniature accomplishments. Despite being a plant biologist, she's a great scientist and is going to do amazing things. She inspires me to try new things with my work and to think about life in new and interesting ways.

The Djuranovic family has a very special place in my heart. Sergej, Slavica, Vas, and Mila are my second family. Their home was always open to me, and they've helped me through so many events. They've encouraged me in my passions and listened to my frustrations. They were always there with a coffee, a beer, or a delicious meal. Celebrating Slava with them is one of my happiest times of the entire year.

When I started graduate school, one of the best pieces of advice that was given to me was to "join a lab where you can thrive." The Andrew lab has truly been the best place for me and my growth as a scientist and a person. Debbie's unwavering enthusiasm and patience are legendary. I was fairly untrained when I joined the Andrew lab and all of the credit for what I've learned goes to her. The other members of the lab – Afshan, Rebecca, SeYeon, Aria, Jessica,

Kyla, Dorothy, Sangjoon, Yim, Bilal, Raj, Mike, and others – have been great coworkers and friends and have contributed many helpful suggestions over the years. In particular, the postdocs in our lab have given me superb advice about how to overcome challenges in graduate school. I'm forever thankful for their input! I would also like to thank members of the Matunis lab for being great neighbors, and especially to Maggie for her advice and friendship.

My friends have been vital to the completion of my thesis work. I've made amazing friends in Baltimore. The epicenter of many of these friendships is a now-closed corner dive bar in Canton. Together with these friends, I've watched many games of football (Here we go, Steelers…) and won many games of trivia. I'd also like to acknowledge the members of my intramural football team, who are intelligent, athletic, and fun people. I would especially like to thank Lisa, Nina, Jess, Gayle, Natalie, and Jon, who have been the greatest friends a pre-doctoral fellow could ask for. A special part of my heart goes to thanking Alex, who is always quick with a smile, who knows how to make me smile, and who would brag about my work to anyone who could listen. His enthusiasm has brought me out of many work-induced funks, and together we've had many adventures with hopefully many, many more to come. My friends from college and home have always been supportive of my endeavors. So to Lauren, Evan, HD, Lauren Huff, Jonelle, KTP, Ali, Jen, and Maureen – thank you!

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

Chapter 1:

CHAPTER 2:

CHAPTER 3:

INTRODUCTION

All organisms begin life as a single cell. How does this single cell provide the basis for an entire plant, worm or human? Within an organism, how does a group of cells become an arm? How does this arm know where and when to form, and how does it "know" that it's not a brain or an eye or a foot? These types of questions fuel the field of developmental biology, and are partially answered by the genetic and transcriptional pathways that govern development.

In the Andrew Lab, we use the *Drosophila* embryonic salivary gland to study the process of development. This simple model simple is specified by the activity of three transcription factors (*Sex Combs Reduced, extradenticle*, and *homothorax*) that impart salivary gland identity to the cells that form the gland placode (Henderson and Andrew, 2000). The salivary gland starts as a flat placode of cells on the surface of the embryo that then invaginate to form the nascent salivary gland tube (Myat et al 2000b). As development progresses, the gland continues to internalize and move dorsally in the embryo. Shortly thereafter, the salivary gland begins to actively migrate posteriorly in the embryo. Throughout this process, the cells of the salivary gland are also undergoing active rearrangement. These rearrangements enable the gland to form a single layer epithelial tube; without them, flat placode would invaginate into something more resembling a cone. Much work has been done by our lab and others to understand the mechanism that govern gland invagination and migration (Myat and Andrew, 2000; Bradley et al 2003; Vining et al 2004; Harris and Beckendorf,2005 Kolesnikov and Beckendorf, 2007), but several key questions remain. For example, what triggers the first cell to invaginate at the start of salivary gland internalization? Moreover, the migration pathways that affect salivary gland migration do not cause fully penetrant defects, suggesting the presence of additional pathways that influence migration.

G-protein coupled receptors (GPCRs) are interesting candidate molecules for answering some of these questions. GPCRs span the plasma membrane seven times, and are able sense

signaling molecules or ligands with the extracellular domains, and trigger signaling cascades through their intracellular domains. GPCRs are involved in migration of neural crest stem cells (Kruger et al 2003), single celled amoeba (Insall et al 1994), and clusters of lymphatic endothelial cells (Valtcheva et al 2013). A former graduate student, Melissa Vining, surveyed the Berkeley Drosophila Genome Project Expression Profile database and found that three GPCRs are expressed in the *Drosophila* embryonic salivary gland. One GPCR, *CrzR*, is a peptide hormone receptor that is not expressed until the end of embryogenesis when the glands have stopped migrating. Therefore, we suspected that this receptor had little to do with sensing signals that drive migration. The two other genes – *Tre1* and *mthl5* —became the basis for my thesis work.

Tre1 is a member of the Orphan family of GPCRs in the fly and is most closely related to Rhodopsin family melatonin receptors in mammals. As its name suggests, the ligand for Tre1 is unknown. Work from other labs showed that *Tre1* directs the migration of single-celled germ cells to the gonad (Kunwar et al 2003). Since *Tre1* is expressed in the salivary gland during its migration, we investigated if *Tre1* can also influence salivary gland migration. Using a previously characterized mutant allele ($Tre1^{\Delta EP5}$), we examined gland migration and morphology and saw no major defects. Closer examination of this allele, which was a small deletion in the 5' UTR of *Tre1*, revealed that it was not a null allele and still expressed normal levels of *Tre1* transcript. Therefore, this allele did not allow us to answer our initial question. To address this, we created a *Tre1* null allele by homologous recombination (heretofore referred to as $Tre1^{KO}$, where the entire open reading frame of *Tre1*was replaced by the *white+* marker gene.

Characterization of the $Tre1^{KO}$ allele showed major defects in germ cell migration. These defects were much more severe than the previously reported phenotypes. Moreover, *Tre1* was initially characterized because the germ cells were trapped in the endoderm (Kunwar et al 2003), but our analysis showed that germs cells in a $Tre1^{KO}$ disperse throughout the entire

embryo. This phenotype similar to that of the reported *scattershot* (*sctt*) allele of *Tre1*, a mutant that lacks an arginine residue crucial for Rhodopsin GPCR signaling (Kamps et al 2010). Moreover, $Tre1^{\Delta EP5}$ had been shown to affect the polarized localization of proteins such as E-Cadherin and Rho1, a phenotype that we did not observe. Instead, our data, along with the *sctt* work, suggests that loss of *Tre1* does not affect the mechanics of migration, but instead affects directed migration.

We then used the new allele to examine salivary gland migration. Unlike the germ cells, salivary gland migration was unaffected by the loss of *Tre1*. The salivary gland migrates as an intact, polarized epithelium, and it receives multiple migratory cues from nearby tissues, such as the mesoderm, central nervous system, and fat body (Harris and Beckendorf, 2005; Vining et al 2005; Kolesnikov and Beckendorf 2007). Loss of a single signaling pathway may not grossly affect salivary gland migration due to redundant cues from multiple tissues. Closer examination of the salivary gland did reveal altered microtubule dynamics in a $Tre1^{KO}$. *Tre1* is known to play a role in microtubule dynamics as it relates to polarized cell division of neuroblasts, and this role appears to be conserved in the salivary gland. We plan to further investigate these changes as they may cause subtle defects in salivary gland migration that have not been previously observed.

The second GPCR I examined is *methuselah like 5* (*mthl5*). The Methuselah family has sixteen members, and only one (*mist*/*mthl1*) has been fully characterized in the *Drosophila* embryo. For decades, it was known that a GPCR was involved in gastrulation (Costa et al, 1994; Parks and Wieschaus, 1991). Both a ligand (*fog*) and the downstream signaling pathway (the Gα *concertina*) were known, but the receptor linking the two remained elusive. Recently, Mist was identified as the GPCR that binds Fog and mediates gastrulation. Because (1) Mist and Mthl5 are closely related, and (2) the cell shape changes that occur during gastrulation are similar to those

that happen during salivary gland invagination, we asked if Mthl5 mediates cell shape changes during the initial steps of salivary gland invagination, and if Fog is also a Mthl5 ligand. We found that Mthl5 is strongly localized to the apical membrane at the early steps of salivary gland invagination. Using the CRISPR/Cas9 system, we created a null allele of *mthl5*. We also made constructs for *mthl5* over-expression and obtained *fog* mutant and over-expression lines. All of these constructs showed similar irregularities following invagination in the salivary gland apical membrane, suggesting an interaction between *fog* and *mthl5*. Given the nature of the irregularities and the early expression of both *mthl5* and *fog* in the salivary gland, we wanted to look more closely at the cellular dynamics early in invagination. Interestingly, cells in the salivary gland placode do not decrease their apical area in a *mthl5* mutant as they do in wild-type glands. This strongly suggests that Mthl5 is binding Fog and triggering the initial cell shape changes that drive invagination. Together, my work investigating *Tre1* and *mthl5* add insight into the role of GPCR signaling in tissue migration.

The final project that I worked on was a departure from the GPCR studies. This work involved the characterization of the Fox (Forkhead box) transcription factor *foxL1*. Fox proteins play a variety of roles in *Drosophila* development (Lee and Frasch). *forkhead* (*fkh*) itself plays a key role in salivary gland migration – in *fkh* mutants, the salivary gland cells do not undergo apical constriction and thus do not invaginate (Myat and Andrew, 2000a). The expression pattern of *foxL1* (previously known as *fd64a*) was first observed by Rika Maruyama, a former post-doctorate fellow in the Andrew lab. Interestingly, the salivary gland migrates directly on top of a small subset of cells expressing *foxL1*. We indentified these cells as Ventral Intersegmental 5 (VIS5), a muscle that is located only in the thoracic region of the embryo. Moreover, I found that VIS5 is not one large muscle as previously described (Bate 1990), but is three independent fibers in each of the three thoracic segments. *foxL1* is expressed in only the

second and third segments, which is where the salivary gland is positioned. Because of the intimate association of the salivary gland and the *foxL1* expressing cells, I wondered if *foxL1* and its targets were influencing salivary gland guidance. I created several lines to study the function of *foxL1*, including a knock-out via homologous recombination, a null allele via CRISPR/Cas9, and an mCherry knock-in that enables us to better characterize the *foxL1* expressing cells. While loss of *foxL1* does not overtly affect salivary gland migration, mis-expression of *foxL1* causes the salivary gland to mis-migrate. Moreover, the same mis-expression causes severe defects in muscle formation. These data suggest that targets of *foxL1* may include signaling molecules that affect how cells signal to one another. I identified to known signaling molecule *sema2a* as a target of *foxl1* in VIS5 of thoracic segment three. Our future plans include performing a microarray with the alleles I have made to gain a better understand of the targets of *foxL1* that are causing salivary gland mis-migration and muscle malformation.

Overall, my thesis work has contributed to a better understanding of how the salivary gland is formed. The characterization of *Tre1* and *mthl5* have broader implications for how GPCRs function during development. Now that I have created null alleles of *Tre1*, *mthl5*, and *foxL1*, our lab can begin to evaluate the role of compensatory signaling in directing salivary gland migration. By using double and triple mutants, we can examine the more nuanced roles that multiple signaling pathways play in directing the migration of an entire tissue.

REFERENCES

Bate, M. (1990) The embryonic development of larval muscles in Drosophila. *Development*, **110**(3),791-804.

Costa, M., Wilson, E. T., and Wieschaus, E. (1994). A Putative Cell Signal Encoded by the folded gastrulation Gene Coordinates Cell Shape Changes during Drosophila Gastrulation, *Cell*, **76**(6), 1075-1089

Harris, K.E. and Beckendorf, S.K. 2007. Different Wnt signals act through the Frizzled and RYK receptors during Drosophila salivary gland migration. *Development* **134**: 2017-2025.

Henderson, K.D., and Andrew, D.J. (2000). Regulation and function of Scr, exd, and hth in the Drosophila salivary gland. *Dev. Biol.* **217**(2), 362-374.

Insall, R.H., Soede, R.D., Schaap, P., and Devreotes, P.N. (1994). Two cAMP receptors activate common signaling pathways in Dictyostelium. *Mol Biol Cell* **6**, 703-711.

Kamps, A.R., Pruitt, M.M., Herriges, J.C., and Coffman, C.R. (2010). An Evolutionarily Conserved Arginine Is Essential for Tre1 G Protein-Coupled Receptor Function During Germ Cell Migration in Drosophila melanogaster. *PLoS ONE* **5**(7): e11839.

Kolesnikov, T. and Beckendorf S.K. (2005). Netrin and Slit guide salivary gland migration. *Developmental Biology* **284**(1): 102-111.

Kruger, G.M., Mosher, J.T., Tsai, Y.H., Yeager, K.J., Iwashita, T., Gariepy, C.E., Morrison S.J. (2003). Temporally distinct requirements for endothelin receptor B in the generation and migration of gut neural crest stem cells. *Neuron* **40**, 917-929.

Kunwar, PS, Sano, H, Renault, AD, Barbosa, V, Fuse, Naoyiku, and Lehmann, R. (2008). Tre1 GPCR initiates germ cell transepithelial migration by regulating *Drosophila melanogaster* E-Cadherin. *JCB* **183**(1): 157-168.

Lee, H. H., and Frasch, M. (2004). Survey of Forkhead Domain Encoding Genes in the Drosophila Genome: Classification and Embryonic Expression Patterns. *Developmental Dynamics*, **229**(2), 357–366.

Myat, M. M., and Andrew, D. J. (2000a). Fork head prevents apoptosis and promotes cell shape change during formation of the Drosophila salivary glands. *Development*, **127**(19), 4217–4226.

Myat, M. M., and Andrew, D. J. (2000b). Organ shape in the Drosophila salivary gland is controlled by regulated, sequential internalization of the primordia. *Development*, **127**(4), 679–691.

Parks, S., and Wieschaus, E. (1991). The Drosophila Gastrulation Encodes a Ga-like Protein Gene concertina encodes a Ga-like Protein. *Cell*, **64**(2), 447–458.

Valtcheva, N., Primorac, A., Juisic, G., Hollmen, M., and Detmar, M. (2013). The orphan adhesion g protein coupled receptor GPR97 regulates migration of lymphatic endothelial cells via the small GTPases RhoA and Cdc42. *JBC* **288**, 35736- 35748.

Vining, M. S., Bradley, P. L., Comeaux, C. A., and Andrew, D. J. (2005). Organ positioning in Drosophila requires complex tissue-tissue interactions. *Developmental Biology*, **287**(1), 19–34.

CHAPTER 1:

Building and specializing epithelial tubular organs: the Drosophila salivary gland as a model system for revealing how epithelial organs are specified, form and specialize

This chapter is modified from : Chung, S-Y., Hanlon, C.D., and Andrew, D.J. (2014). Building and specializing epithelial tubular organs: the Drosophila salivary gland as a model system for revealing how epithelial organs are specified, form and specialize. *Wiley Interdiscip Rev Dev Biol*. 3: 281-300.

ABSTRACT

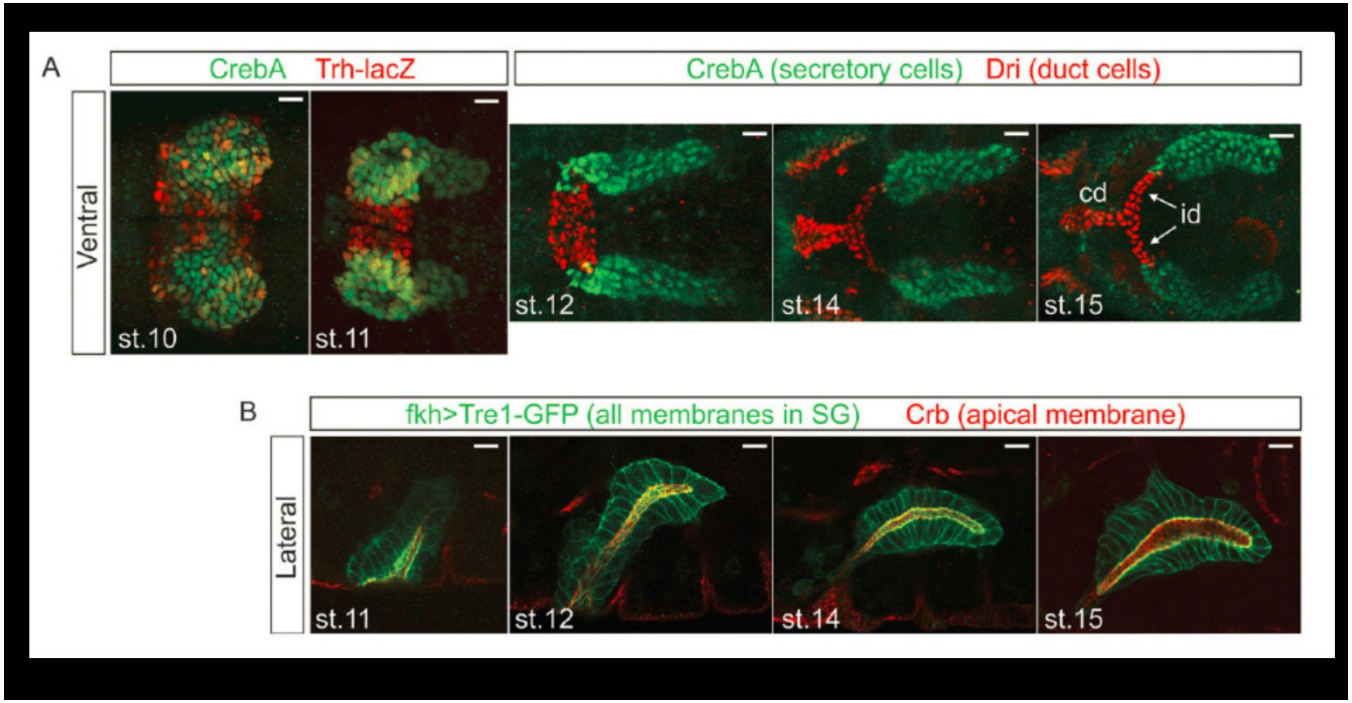
The past two decades have witnessed incredible progress toward understanding the genetic and cellular mechanisms of organogenesis. Among the organs that have provided key insight into how patterning information is integrated to specify and build functional body parts is the Drosophila salivary gland, a relatively simple epithelial organ specialized for the synthesis and secretion of high levels of protein. Here, we discuss what the past couple of decades of research have revealed about organ specification, development, specialization and death, and what general principles emerge from these studies.

INTRODUCTION

Drosophila has proven to be an ideal model system for revealing the molecular and cellular underpinnings of organ development. In particular, the salivary gland (SG) has been crucial for studying how polarized epithelial organs form and specialize. The SG starts out as two plates of approximately one hundred fifty cells each on the ventral surface of the embryo (**Figure 1**). Following internalization and initial tube formation, the fully polarized SG elongates and actively migrates to its final position in the embryo. Each of the secretory glands connects to an individual duct. Individual ducts connect to one another in a central common duct, which attaches to the mouthparts. The SG that forms during embryonic development persists through larval life to the beginning of pupation, when it produces a final burst of secretion before being destroyed during metamorphosis. Subsequently, the imaginal ring cells, which are found between the SG duct and secretory cells and that proliferate during larval and pupal stages, form the adult gland.

As a model system, the SG has been a valuable platform for addressing questions of cell fate specification and maintenance, and for learning how cells coordinate their activities to build an organ of the right size, shape and position in the animal. The SG has been excellent in revealing how the sequential deployment of gene expression programs controls all aspects of epithelial tube form and function. Morphogenetic processes, including invagination, tube elongation and migration, continue to be studied in the SG, and new molecules driving these events are being uncovered. The SG has also been used extensively to study programmed cell death in the context of a living organism. In the following review, we discuss our current understanding of SG formation, maintenance and death, and how findings from these studies have provided insight into more general aspects of organ development and gene function.

Figure 1. Confocal images of the embryonic SG. (A) Ventral views of the SG stained with nuclear markers. The secretory portion of the SG forms from two placodes of cells on the surface of the embryo, with the duct precursors located between the two secretory placodes (st 10). At this stage, the expression of duct (red) versus secretory (green) markers is not so clear. During st 11, the gland invaginates into the embryo. At this stage, the distinction between duct (red) and secretory markers (green) is more evident. As development progresses (st 12-15), the secretory tubes elongate, and the individual duct (id) and common ducts (cd) form. (B) Lateral views of SG during elongation and migration. All membranes are marked in green, with the apical membrane specifically marked in red. Following invagination, the SG moves dorsally (st 11) and then turns (st 12) and migrates posteriorly. Posterior migration continues (st 14-16) until the SG reaches its final resting place. Throughout this dynamic process, the SG migrates as an intact fully polarized tissue.


A
CrebA Trh-lacZ
CrebA (secretory cells) Dri (duct cells)
Ventral
st.10
st.11
st.12
st.14
cd
id
st.15
B
fkh>Tre1-GFP (all membranes in SG)
Crb (apical membrane)
Lateral
st.11
st.12
st.14
st.15

SPECIFICATION OF SALIVARY GLANDS

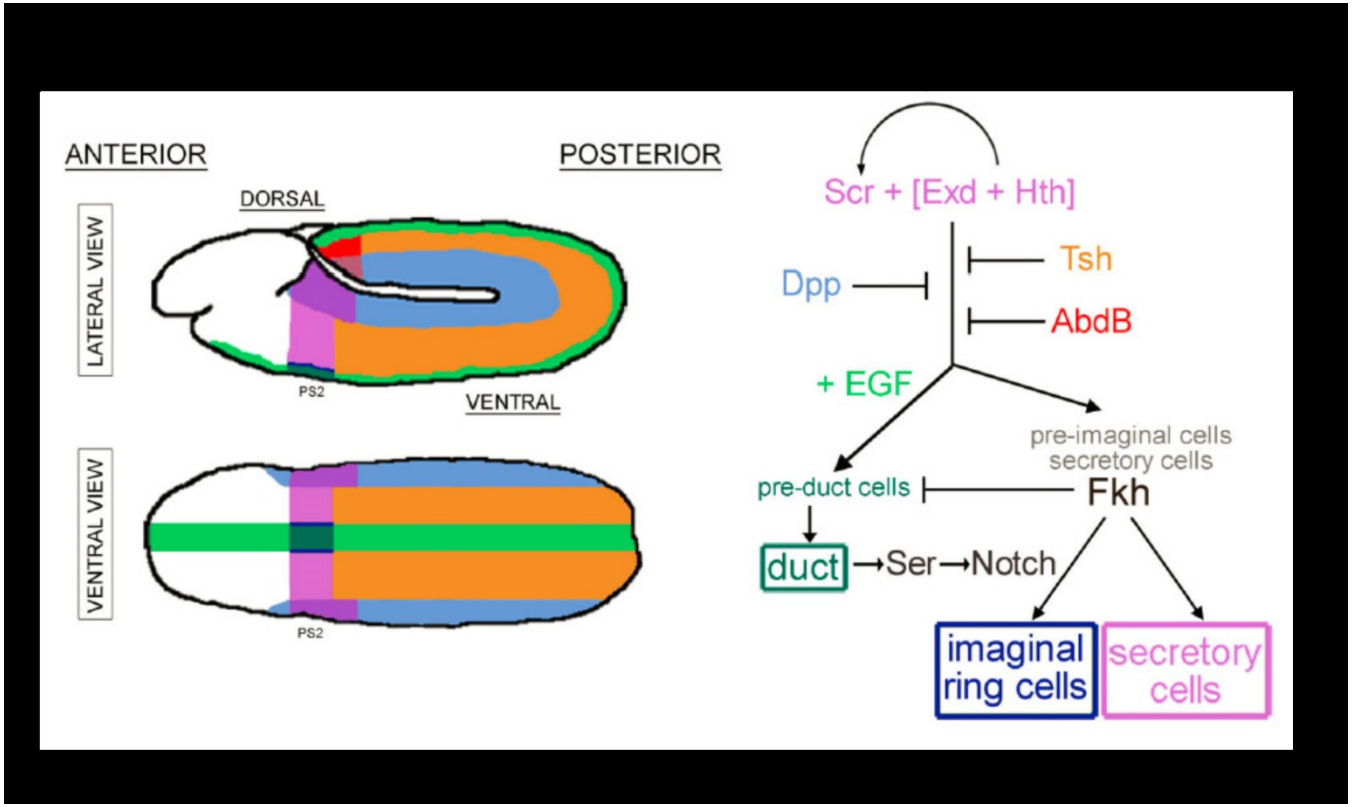
SG specification and the distinction between secretory, duct and imaginal ring (pre-adult) cell populations occur through the integration of anterior-posterior patterning information, largely mediated by HOX genes, and the dorsal-ventral patterning system, specifically Dpp-signaling in dorsal cells and EGF-signaling along the ventral midline. Since SG precursors stop dividing once specified, the process of specification determines both where the primordia will arise and the final number of cells in the fully formed tissue.

Anterior-posterior patterning genes and SG specification

Scr, Exd and Hth: positive determinants of SG fates

SG specification requires positive input from three transcription factors: Sex combs reduced (Scr), Extradenticle (Exd) and Homothorax (Hth) (**Figure 2**)[1-3]. SGs form from the ventral ectodermal cells of parasegment 2 (PS2), which express Scr, the only spatially limited component required to activate SG formation. Exd, a TALE (*t*hree *a*mino acid *l*oop *e*xtension) homeodomain protein that is expressed throughout the embryo[3], is also required for SG formation. Scr and Exd bind each other and bind DNA. The physical association of Exd with Scr distorts the N-terminus of the Scr homeodomain, allowing it to fit into the relatively narrow minor groove of its target DNA sites[4]. Exd activity is limited by regulated nuclear entry, mediated through binding to its essential cofactor Hth[5]. Hth, another broadly-expressed TALE protein with highly conserved mammalian orthologues (MEIS1,2,3 and PREP1,2 proteins) is also absolutely required for SG formation[2]. Whether Hth is directly involved in the binding of the Scr-Exd complex to target sequences is not clear, but the homeodomain of Hth, which is present in only one of two alternative splice forms, is not required for Scr-Exd dependent activation of the single SG target reporter that has been directly tested[6]. Early expression of Scr in SG precursors

Figure 2. SGs are specified by the integration of patterning information along both major body axes. Scr (purple in cartoon), a Hox protein expressed in PS2 and dorsal cells of PS3, is the only spatially-regulated activator of SG cell fates. The more globally expressed Exd and Hth Tale-homeodomain proteins are also required for SG formation and these proteins function at multiple levels. Tsh (brown in cartoon), expressed in PS3-13, and AbdB (red in cartoon), expressed in PS14, block SG activation in the trunk and abdomen. SG formation is limited to the ventral cells of PS2 by Dpp signaling (blue in cartoon) in the dorsal cells. EGF-signaling (green in cartoon) along the ventral midline specifies the duct cell fate by blocking expression of Fkh. In turn, Fkh plays a major role in maintaining the secretory cell fate by regulating itself as well as multiple other secretory-specific genes and by blocking expression of duct-specific genes. Ser, which is expressed in duct cells, signals adjacent Notch expressing cells in the common secretory/imaginal ring cell primordia to become imaginal ring cells (dark blue in diagram) – the precursors to the adult SG.


ANTERIOR
POSTERIOR
DORSAL
LATERAL VIEW
PS2
VENTRAL
VENTRAL VIEW
PS2
Scr + [Exd + Hth]
Dpp
Tsh
AbdB
+ EGF
pre-imaginal cells
secretory cells
pre-duct cells
Fkh
duct
Ser
Notch
imaginal ring cells
secretory cells

requires both Exd and Hth, revealing that TALE proteins function at multiple levels in SG specification[2]. Most other Hox proteins also require Exd and Hth for target gene regulation[3,7]. One exception is Abdominal-B (Abd-B), which instead represses Exd and Hth expression to mediate specification of another embryonic organ in Drosophila, the posterior spiracle[8,9].

Ectopic expression of Scr driven by heat shock or by ubiquitous Gal4 causes ectopic expression of many SG markers in all segments at early stages[1,4,10,11]; expression of SG markers persists, however, only in PS2 and the two parasegments more anterior, PS0 and PS1. Expression of SG markers in posterior segments is transient, disappearing completely by mid-embryogenesis[11]. Moreover, SG marker-expressing cells in posterior regions do not invaginate to form SGs – which does occur in the more anterior regions. Persistent expression of SG markers induced by global expression of Scr is only observed in PS3-13 in the complete absence of *teashirt* (*tsh*) function (see below)[10].

Posterior Hox genes and Tsh: Negative regulators of SG fates

SG formation is further defined by negative regulation from two transcription factors. The major block to SG activation in PS3-13 is Teashirt, a zinc finger containing transcription factor (**Figure 2**)[10,12]. In the absence of Tsh, Scr early expression expands to PS3 and SGs form in PS3. A role for Tsh in blocking Scr-induced SG formation fits well with the previously described role for Tsh in distinguishing trunk from head[13,14]. Tsh directly regulates at least one SG target gene [12] and Tsh has been shown to physically contact Scr [15]; the exact mechanisms whereby Tsh prevents Scr's SG-inducing activity, however, remain unclear. Nonetheless, Tsh appears to function at two levels: repression of early Scr expression in the ventral cells of PS3 and repression of Scr's SG-inducing activities when that transcriptional regulation is overridden [10].

In PS14, Scr-induced SG gene activation is blocked by another Hox gene, *Abdominal B* (*Abd-B*) (**Figure 2**)[10]. The block by Abd-B is not absolute and may be due to "posterior prevalence", a phenomenon wherein more posteriorly expressed Hox proteins block the activities of more anteriorly expressed Hox proteins[16-21]. Recent work suggests that posterior prevalence occurs because posteriorly expressed Hox proteins compete more successfully for the shared essential cofactor Exd[6]. Alternatively, given recent findings that Abd-B shuts off expression of Exd and Hth, the block by Abd-B could be due to the absence of these essential SG-inducing cofactors in PS14[8].

Dorsal-ventral patterning genes further refine SG coordinates

Although Scr, Exd and Hth drive SG formation, not all cells that express these transcription factors become SGs; SGs form only from ventral ectodermal cells. Thus, dorsal-ventral patterning information also feeds into the system. Loss of *dorsal* (*dl*), a major early determinant of ventral cell fates, results in a complete loss of SG marker expression[1]. Dl promotes SG development by blocking ventral expression of the gene encoding the BMP ligand Decapentaplegic (Dpp), which is normally expressed in only dorsal cells. In turn, Dpp signaling blocks SG formation (**Figure 2**); loss of any one component in the Dpp signaling pathway results in the dorsal expansion of all tested SG markers[1,22,23].

Maintaining the SG fate

Although Scr, Exd and Hth are absolutely required for SG formation, their expression/nuclear localization disappears from the SG shortly after the onset of morphogenesis[2]. So, how is cell fate maintained in this tissue? Among the genes activated by Scr, Exd and Hth is the winged

helix transcription factor Fork head (Fkh), a key player in maintaining SG fates (**Figure 3**). Fkh, the single FoxA family protein in flies, maintains its own expression[24] and the expression of two other transcription factor genes that are highly expressed in the secretory cells – *CrebA*, which encodes a bZip transcription factor related to the Creb3/Creb3L family of mammalian proteins, and Sage, which encodes a less well conserved bHLH transcription factor distantly related to several mammalian proteins[25-27]. *fkh* affects the expression, either directly or indirectly, of about 60% of all SG genes, including those it represses in the SG duct (see below), as well as downstream targets of CrebA and Sage[11,28]. As will be discussed later, sustained expression of Fkh, Sage and CrebA is critical for SG function[26-29]. Fkh, Sage and another downstream transcription factor Senseless (Sens) are also required for secretory cell survival; embryos mutant for any one of these three genes undergo extensive apoptotic SG cell death following elevated expression of the apoptotic activator genes, *reaper* (*rpr*) and *head involution defective* (*hid*)[25,28,30]. Moreover, Fkh is essential for SG morphogenesis[25].

Duct versus secretory fate specification

Ducts form from approximately 25-30 cells on each side of PS2 between the ventral midline and the more laterally positioned SG placodes. Once the SG has internalized, duct and secretory cells can be distinguished at the molecular level since expression of several genes is restricted to only one of these domains (**Figure 1**). For example, *fkh* and its many transcriptional targets, are expressed to high levels in the gland cells and levels are either reduced or absent in duct cells. In contrast, expression of several other genes, including *Serrate* (*Ser*)[31 32], *breathless* (*btl*)[33] and *dead ringer* (*dri*)[34], is duct-specific. The secretory versus duct cell distinction is not so clear prior to invagination. Although levels of secretory gene expression in the cells flanking the midline are generally lower, there is considerable cell-to-cell variation, variation that disappears as

Figure 3. The secretory specific genes directly activated by Scr, Exd and Hth include several transcription factors – CrebA, Fork head (Fkh), Sage and Huckebein – as well as a splicing factor – Pasilla. The early expressed genes both maintain and implement the secretory cell fate decision. Pasilla and CrebA function to increase secretory capacity in the professional secretory cells of the SG. Fkh maintains its own expression as well as expression of CrebA and Sage. Fkh also controls morphogenesis and in collaboration with Sage and their downstream target Senseless, activates expression of SG-specific genes and represses the apoptotic genes *reaper* and *hid*, keeping the SGs alive. Huckebein mediates tube elongation.

Scr + Exd + Hth
Huckebein
Sage
Fkh
CrebA
Pasilla
Senseless
tube elongation
survival (rpr, hid) tissue specific genes (secreted, TM proteins, enzymes)
morphogenesis
secretory capacity

morphogenesis proceeds (**Figure 1**). The changes in secretory versus duct cell marker expression may reflect the graded distribution of the signal(s) that induce(s) specific cell fates (e.g. EGF – see below) followed by boundary sharpening through repression and/or activation of gene expression by the transcription factors expressed to higher levels in one versus the other cell type.

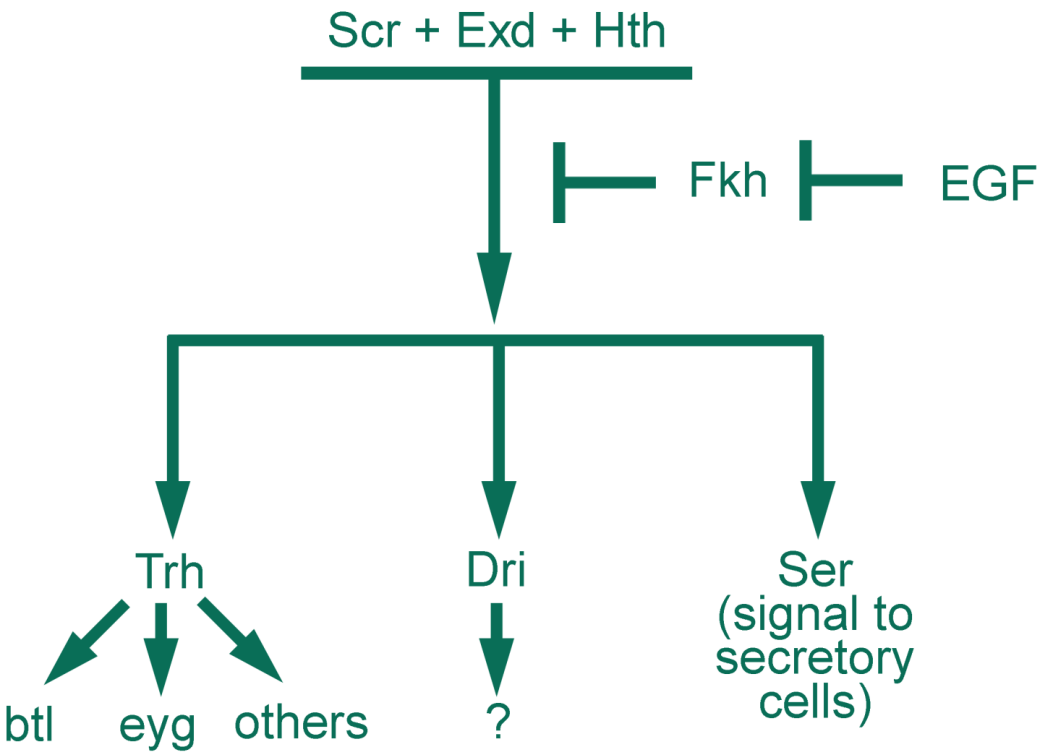
Dorsal specifies duct cell fates through activation of EGF signaling

Dl plays an instrumental role in distinguishing SG duct versus secretory cell fates (**Figure 4**). Dl (through both direct and indirect mechanisms) activates expression of the bHLH transcription factor Single minded (Sim) along the ventral midline[35]. In turn, Sim activates expression of *rhomboid* (*rho*)[36], which encodes the spatially limited component required for Epidermal Growth Factor (EGF) signaling[37]. Sim, Rho and the EGF ligand Spitz are all required for duct specification; loss-of-function mutations in any one of these genes results in a ventral expansion of secretory cell markers at the expense of duct cell markers[1,38].

EGF signaling specifies duct cell fates by repression of fkh

The boundary between gland and duct cells is determined by the combination of two negatively regulated steps; the EGF signaling pathway represses *fkh* expression in the duct cells and Fkh represses expression of duct-specific genes in the gland cells[38,39]. Trachealess (Trh), a basic helix loop helix (bHLH)-PAS transcription factor, is initially expressed throughout the entire SG primordia[22]. As with other duct genes, *trh* expression is repressed by Fkh, remaining on in only duct cells[22,38]. Early experiments suggested that loss of *trh* resulted in a loss of expression of all other tested duct markers[38]; Trh appears, however, to affect expression of only a subset of duct genes, in many cases simply boosting their expression levels[39]. Thus, EGF-dependent duct cell

Figure 4. The duct is specified by Scr, Exd and Hth in combination with EGF signaling. EGF signaling blocks expression of Fkh in the most ventral cells of PS2, allowing the expression of duct-specific genes such as Trh and its downstream targets, *btl*, *eyg* and, presumably, other genes. The absence of Fkh expression in the ventral cells also allows expression of two other duct-specific factors: Dri and Ser. Ser in duct cells is important for establishing the imaginal ring cell fate.


Scr + Exd + Hth
Fkh
EGF
Trh
Dri
Ser
(signal to
secretory
cells)
btl
eyg
others
?

specification appears to be mediated largely by repression of *fkh* expression in future duct cells (**Figure 4**).

Notch signaling specifies the adult SG primordia

Notch signaling through Ser, one of two Drosophila Notch ligands, specifies the imaginal ring cells (**Figure 2**)[39]. *Ser* expression is duct specific[31,32], whereas Notch is transiently upregulated in secretory cells[40]. *Ser* loss results in the absence of imaginal ring cells[39]. Correspondingly, a mutant allele of *Notch* that specifically affects its response to Ser, but not to Delta, the other Notch ligand in flies, also results in a loss of the imaginal ring cells[41]. Although it is clear that imaginal ring cells are specified during embryogenesis[39], whether this happens prior to or during tube morphogenesis is unknown. Roles for Notch signaling in specifying distinct cell types within developing organs has also been documented in the vertebrate vasculature[42], pituitary gland[43], pancreas[44] and central nervous system[45].

CONSTRUCTION OF THE SECRETORY TUBES

Once specified, the SG primordial cells undergo morphological changes and internalize to form epithelial tubes. Since no further cell division or cell death occurs during SG differentiation, all subsequent changes take place within and between pre-existing cells. Polarity is maintained throughout the entire process of tube morphogenesis and, in the fully internalized SG tubes, the apical domain forms the luminal surface and the basal domain contacts surrounding tissues (**Figure 1**).

Regulated sequential secretory cell internalization

The first SG cells to internalize are located in the dorsal-posterior regions of the SG placodes, plate-like structures of columnar epithelial cells that form shortly after Scr expression is first observed in the primordia. The dorsal-posterior cells undergo apical constriction, a process whereby nuclei move to the basal domain and the apical domain constricts to create the pyramidal shaped cells thought to drive tube internalization[46]. Subsequently, neighboring cells invaginate to create nascent epithelial tubes, with their newly formed lumens contiguous with the apical surface of the SG cells still on the surface. Through a series of less well-characterized shape changes and cell rearrangement, the remaining placode cells internalize to form elongated fully internalized epithelial tubes.

Fkh is required for secretory cell internalization

Fkh plays a major role in SG morphogenesis; the SG primordia in *fkh* mutant embryos rescued from cell death do not undergo apical constriction and completely fail to internalize. Basal movement of nuclei, however, is unaffected by *fkh* loss, indicating that the two processes – apical constriction and basal nuclear movement – are separable[25]. Fkh, as a transcription factor,

likely controls SG invagination (and other processes) indirectly, through its downstream targets; the Fkh targets that mediate internalization, however, remain to be identified and/or characterized.

Rho-GTPases and the actinomyosin cytoskeleton mediate secretory gland internalization

Several molecules that directly affect SG internalization have been discovered, many of which are known to be more generally involved in cell shape changes (**Table 1**). Mutations in *folded gastrulation* (*fog*), a secreted ligand for the G protein-coupled receptor Mist[47], and RhoGEF2, a Rho GTPase exchange factor required for invagination of the ventral furrow[48,49], result in partial failure of SG internalization[49,50]. RhoGEF2 affects apical constriction by regulating apical accumulation of Spaghetti squash (Sqh), the Myosin regulatory light chain[49]. Embryos mutant for 18 wheeler (18W), a Toll-like receptor protein, and two Rho-GAPs, RhoGAP5A and RhoGAP88C/Crossveinless-c (CV-C), also show delays and/or partial failure of SG internalization, further supporting the idea that Rho signaling is critical[50]. Indeed, both Rho1 mutants and embryos with SG-specific expression of a dominant-negative Rho1 construct showed partial defects in SG invagination[51]. Rho1 regulates SG internalization by two mechanisms: 1) upregulation and apical localization of transcripts for *crumbs* (*crb*), which encodes an apical membrane protein necessary for the establishment and maintenance of apical-basal polarity and for apical membrane expansion[52-55], and 2) induced apical constriction and cell shape changes mediated by Rho-kinase (Rok)[51].

As with other morphogenetic processes, cytoskeletal events associated with cell shape changes are key for SG tube formation. Indeed, a prominent multi-cellular Myosin II cable forms around the SG placode prior to invagination. This cable is maintained throughout the process of SG

Table 1: Genes implicated in salivary gland (SG) morphogenesis with their currently understood roles and interactions

Process: Secretory cell invagination	
Fog	Ligand for GPCR Mist in mesoderm Coordinates invagination
Rho1	GTPase Increases Crb levels Maintains Crb, aPKC, and Stardust apical localization
Crb	Apical membrane protein Establishes and maintains apical/basal polarity Along with aPKC, prevents Rok accumulation and MyoII cable formation in SG
aPKC	Along with Crb, negatively regulates Rok and MyoII
Rok	Rho kinase Induces cell-shape changes and apical constriction Positive regulator of MyoII
RhoGEF2	Rho guanine nucleotide exchange factor Regulates apical constriction via Sqh localization
Sqh	Myosin regulatory light chain Creates intracellular myosin network
RhoGAP88C	Basolaterally localized GTPase activating protein Coordinates invagination
18-Wheeler	Toll-like receptor Component of Rho pathway Coordinates invagination
Tec29/Btk29A	Non-receptor tyrosine kinase Affects F- and G- actin dynamics Coordinates invagination and migration
Gyc76C	Guanylyl cyclase Upstream of DG1 Regulates Talin and laminin localization
DG1	cGMP-dependent kinase 1 Downstream of Gyc76C Regulates Talin and laminin localization

Process: Tube elongation via cell elongation	
Crb	Transmembrane protein Mediates apical membrane expansion
Rib	Transcription factor Upstream of Crb Regulates moesin activity (indirect)
Lolal	Mediates nuclear localization of Rib in the SG (and other ectoderm)
Moesin	ERM family member; actin linker Increases apical stiffness when phosphorylated
Cdc42	Rho GTPase; activates Pak1
Pak1	Serine-threonine kinase Regulates lateral E-Cad endocytosis through Merlin, Dynamin, and Rab5
E-Cadherin	Cell-cell adhesion Mediates cell shape changes in the SG [e.g. expansion along the apical P/D axis]

Process: Tube elongation via cell rearrangement	
Gon1	Genetically interacts with Cad99C
Cad99C	Apically localized protocadherin Influences apical membrane release from ECM and consequent cell rearrangements

internalization and cinches up as more cells are internalized, suggesting that tension created by this cable provides a motive force driving internalization[56]. The myosin cable forms at the interface between peripheral SG cells and their immediate non-SG neighbors through downregulation of Rok. The high levels of Crb and atypical protein kinase C (aPKC) found within SG cells (and not in neighboring non-SG cells) negatively regulate Rok accumulation, preventing the formation of the Myosin II cable within the placodes[56]. Actin reorganization is also essential for timely invagination of the SG. Null mutations in Tec29/Btk29A, a member of the Tec family of non-receptor tyrosine kinases, cause a delay in SG invagination due, in part, to a shift in the equilibrium between F- and G-actin[57]. Although most mutants with invagination defects appear to affect the apical surface, a recent study has revealed that Guanylyl cyclase at 76C (Gyc76C) and its downstream cGMP-dependent kinase 1 (DG1) affect invagination, collective migration and SG lumen shape partly by regulating localization of Talin and the laminin matrix surrounding the basal surface of the SG[58].

ELONGATION OF THE SECRETORY TUBES

As the SG primordia internalize and the resulting tube moves to its final correct position in the embryo, the tube elongates through both cell shape change and cell rearrangement.

Tube elongation by cell elongation

Hkb regulates polarized growth and delivery of apical membrane

Polarized growth and delivery of apical membranes is critical for tube elongation[59]. In embryos mutant for *huckebein* (*hkb*), which encodes a zinc finger transcription factor[60], the SG cells internalize but almost completely fail to elongate, resulting in small 'puck-shaped' SGs with very little apical surface area[46,59]. Hkb controls SG apical expansion through increased translation and/or stabilization of Crb[52-55], and increased transcription of *klarsicht* (*klar*), which encodes a putative regulator of the dynein ATPase that mediates microtubule-dependent vesicle transport to the apical surface[61,62]. Thus, Hkb facilities tube elongation through both Crb-mediated apical membrane expansion and Klar-driven apical targeting of membrane vesicles (**Table 1**).

Ribbon reduces apical stiffness to facilitate membrane expansion

Ribbon (Rib), a BTB-containing transcription factor, modulates apical membrane expansion to elongate the SG, and two other tubes, the trachea and Malpighian tubules[63-65]. *rib* mutant SGs achieve only 60% of the WT lumen length and live imaging studies indicate that *rib* mutant SGs elongate more slowly than WT[66]. Rib interacts with Lola like, another BTB-domain containing protein required for robust nuclear localization of Rib, to upregulate *crb* transcription and to downregulate the activity of Moesin (Moe), a protein that cross-links the apical membrane to the apical cytoskeleton[65,67,68]. Genetic and mechanical analyses suggest that the increased apical stiffness caused by increased Moe activity (increased phosphorylated Moe) is a major

contributor to the *rib* mutant phenotype[65,66]. Thus, tube elongation requires not only the generation of sufficient apical membrane for expansion but also modulation of the mechanical properties of the apical domain (**Table 1**).

Tube elongation by cell rearrangement

Rho1 controls tube length by cell elongation and rearrangement

To achieve fully elongated tubes, the SG also undergoes cell rearrangement. This process simultaneously reduces the number of cells in circumference and increases the number of cells along the length of the tube. Some molecules, such as the Rho1 GTPase, control SG lumen size by regulating both cell rearrangement and cell shape. Proximal SG cells of Rho1 mutant embryos fail to rearrange and the apical domains do not elongate fully[69]. SG-specific knockdown of Rok by RNAi causes similar defects as observed in Rho1 mutants, suggesting that Rho1 affects tube architecture, at least in part, through Rok regulation of myosin mechanics[69]. Rho1 also appears to work with Rib to limit apical phosphorylated Moe, suggesting more direct effects on the apical actin cytoskeleton, an idea supported by the observed changes in F-actin distribution[69]. Interestingly, cell rearrangements mediated by Rho, Rock and Myosin II also drive tube elongation in the vertebrate gut[70], suggesting a conserved role for Rho1 in cell rearrangement.

Reduced cell-cell adhesion facilitates cell rearrangement

Allowing cells to rearrange while maintaining polarity requires tight regulation of the junctional complexes that hold epithelial cells together. Since septate junctions (structures equivalent to vertebrate tight junctions) do not fully mature until late embryogenesis[71], the major junction requiring modulation during SG cell rearrangement is the adherens junction (AJ). Regulation of cell rearrangement in the SG appears to be through Rac1 modulation of the AJ protein E-

Cadherin (E-Cad)[72]. Too little Rac1 – through either mutations in multiple Rac1 genes [Rac1, Rac2 and Mig 2-like] or through SG expression of a dominant-negative Rac1 construct – results in increased E-Cad (and another AJ protein – catenin) at the AJs. The increased pools of AJ-localized E-Cad blocks cell rearrangement. Consequently, the distal gland, which likely forms without significant cell rearrangement, forms relatively normally, but many of the proximal SG cells, which must rearrange during internalization, remain on the embryo surface. As the distal cells continue to migrate posteriorly in the *rac1* mutants, the gland often breaks, resulting in multiple small glands surrounding separate lumens. Correspondingly, excessive Rac activity results in reduced E-Cad at the AJs and a corresponding loss of adhesion – SG cells disperse and eventually die. The dispersion phenotype driven by SG expression of constitutively-active Rac can be rescued simply by overexpressing E-Cad, supporting the idea that Rac affects cell rearrangement largely through the localization of E-Cad. Rac1 appears to modulate E-Cad pools in the different membrane domains by regulated endocytosis, since altering the endocytic pathway also alters the Rac1 phenotypic outcomes[72].

Not surprisingly, E-Cad turnover is also key to the proximal-distal elongation of individual SG cells. In this case, endocytic turnover of E-Cad is differentially regulated along the axis of polarity by p21-activated kinase 1 (Pak1). Pak1 belongs to a serine-threonine kinase family that in other systems binds and is activated by Cdc42 and/or Rac to regulate diverse biological processes[73,74]. SG expression of dominant-negative Cdc42 and/or loss of Pak1 results in a complete loss of lateral membrane pools of E-Cad and a widening of cells along the dorsal-ventral axis. Correspondingly, high-level expression of activated Pak1 depletes the apical pools of E-Cad, resulting in the loss of the single shared apical lumen and the appearance of multiple intercellular lumena. The intercellular lumena arise between cells in the lateral domains from

endocytosed E-Cad containing vesicles[75]. Pak1 activity in the SG depends on Rab5, Dynamin and the ERM protein Merlin, a substrate of Pak1[75,76].

Just as cells have to release their attachment to their neighbors to rearrange, they also have to release their attachment to the matrix[77]. SGs mutant for *AdamTS-A*, which encodes an apically-targeted and secreted zinc metalloproteinase, have highly irregular luminal surfaces due to a failure of the cells to easily rearrange during tube elongation and posterior migration. *AdamTS-A* null mutants also have over-stretched apical domains in the distal-most SG cells and show increased accumulation of apical actin, suggesting that these cells are under increased tension. Further support for a role of AdamTS-A in releasing the apical cell surface from the apical matrix emerges from the finding that null mutations in *Cadherin99C* (*Cad99C*) rescue the apical irregularities associated with AdamTS-A loss[77]. Cad99C, an atypical cadherin with a large extracellular domain, localizes to the apical surface of SG cells and other epithelia[78,79 80]. Through its attachment to (currently unidentified) apically secreted proteins, Cad99C is proposed to resist and balance the forces driving tube elongation[80]. Thus, associations between the apical surface and apical matrix are important in fine-tuning overall SG tube shape.

POSITIONING THE SALIVARY GLANDS

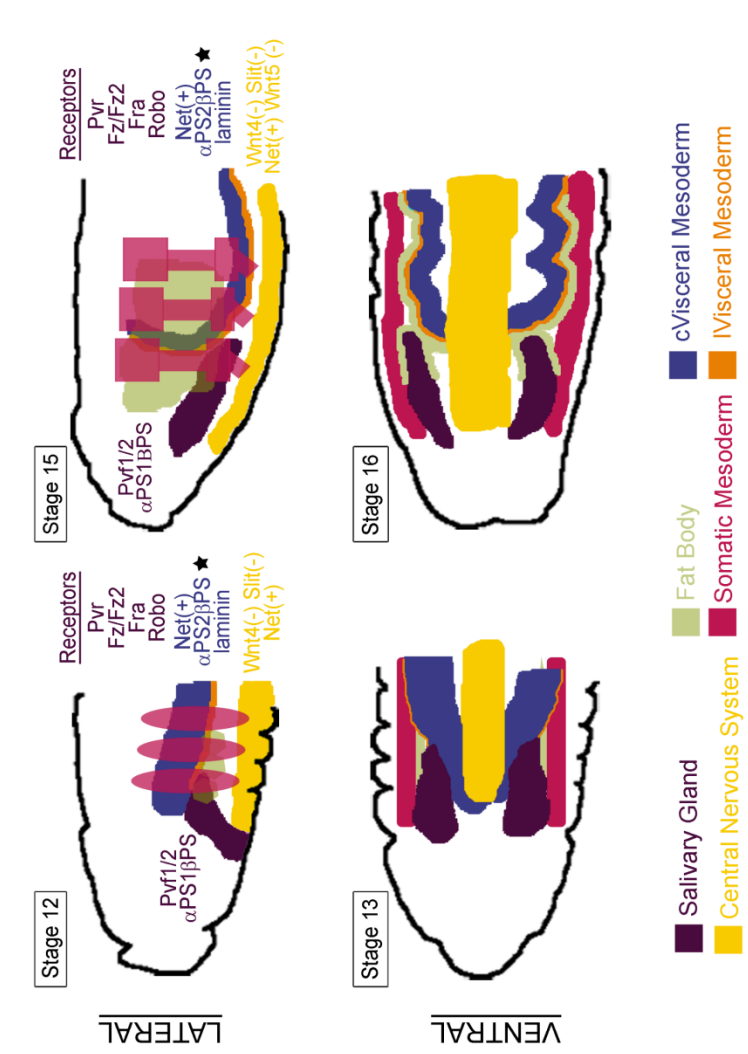
To attain its final correct position in the embryo, the SG actively migrates in direct contact with several other tissues. Whereas some of the contacting tissues may serve only as tracks or barriers to migration, others also provide guidance cues to either attract or repel the gland.

Surrounding tissues promote SG movement

The visceral mesoderm provides a track for SG migration

The internalization process positions the secretory tubes in an approximate dorsal-posterior orientation, with the distal most cells directly contacting the dorsally-positioned visceral mesoderm (VM) (**Figure 5**). As the secretory cells sequentially contact this tissue, they turn and migrate along it to eventually arrive at their final position, with the long axis of the tube aligned with the long axis of the embryo. In mutants where the VM is discontinuous, the SG will often continue to move dorsally instead of turning posteriorly, although the gland still appears to maximize contact with whatever VM tissues remain in these mutants[81]. Posterior migration of the SG absolutely requires integrin expression both in the SG, which expresses □PS1□PS, and in the VM, which expresses not only □PS2□PS, and also an essential laminin that is bound by both integrins[77,81]. With loss of any of these molecules, the SG completely fails to migrate; SG tubes still elongate, however, resulting in buckled, U-shaped tubes. These findings suggest that the VM provides a suitable substrate for SG migration. At later stages, contact between the SG and VM is abrogated by the ingression of caudal mesoderm cells between the gland and VM. Failure of this population to migrate results in the continued attachment of the SG to the VM and, during the process of head involution, this attachment stretches the glands to about twice their normal length[82].

Figure 5. The SG contacts or comes close to several tissues as it migrates to its correct final position in the embryo, including the circular (c) visceral mesoderm (cVM), the longitudinal (l) visceral mesoderm (lVM), the fat body, the somatic musculature and the central nervous system. The cVM provides a suitable substrate for posterior SG migration through the expression of PS2PS integrin that binds a secreted laminin also expressed in the cVM. The SG expresses αPS1βPS integrin, which also binds the secreted laminin. Both the integrins and laminin are essential for posterior migration (starred). The lVM migrates between the SG and the cVM to detach these two cell types. The SG also expresses several receptor genes, which allow it to properly navigate to its final correct position in response to local sources of the corresponding ligands. In turn, the SG is likely to also provide cues for the migration of other cell types in the embryo. For example, the fat body migrates over specific parts of the SG at late embryonic stages.


Stage 12
Stage 13
Stage 15
Stage 16
Receptors
Pvr
Fz/Fz2
Fra
Robo
Net(+)
αPS2βPS
laminin
Pvf1/2
αPS1βPS
Wnt4(-) Slit(-)
Net(+)
Wnt4(-) Slit(-)
Net(+) Wnt5 (-)
LATERAL
VENTRAL
Salivary Gland
Central Nervous System
Fat Body
cVisceral Mesoderm
Somatic Mesoderm
lVisceral Mesoderm

Other tissues also provide SG migration cues

The SG either directly contacts or comes near several other tissues during migration, including the somatic muscle, the fat body, the central nervous system (CNS) and the gastric caeca, long tubular extensions of the midgut that emerge during later embryonic stages (**Figure 5**)[82]. Studies suggest that most, perhaps all, of these tissues provide cues that direct SG movement. Several guidance cues have been discovered that seem to function either to attract or repel the migrating SG.

Molecular guidance

Like most migrating tissues, the SG responds to a range of guidance cues during its posterior migration. The Netrin/Frazzled (Fra), Slit/Robo, and Wnt/Derailed (Drl)/Frizzled (Fz) pathways have all been shown to influence SG migration[83,84] (**Figure 5**). The Netrin ligands are expressed in both the VM and CNS, whereas their receptor Fra is expressed in the SG. Loss-of-function mutations in the *netrin* genes or in *fra* cause a mild migration defect of the glands curving away from the midline[84]. Overexpression of NetrinB (NetB) causes more pronounced mis-migration of the SG towards the source of expression, suggesting that NetB acts as an attractant for the *fra*-expressing SGs. Conversely, Slit acts as a strong SG repellent. Loss of Robo1 and Robo2, receptors for Slit, which is expressed in midline cells, cause the SG to mis-migrate towards the midline[84]. Wnt signaling also acts to repel the migrating SG[83]. Wnt4 and Wnt5 are expressed in the CNS; their respective receptors Fz/Fz2 and Drl are expressed in the SG. Loss of any of these factors causes the SG to curve towards the CNS. The PDGF/VEGF pathway has been implicated in SG migration: mutations in both the receptor and two of the ligands result in ventrally curved SGs[83]. It remains unclear, however, whether this pathway affects migration or some other aspect of SG morphology. Additional cues, provided by other cell types, will likely be discovered

to influence final SG placement; disrupting these pathways, however, may lead to relatively subtle defects in SG placement given the number of tissues providing guidance cues. How an intact epithelial tissue, such as the SG, integrates signals from multiple different sources to control its movement remains to be discovered.

FORMATION OF THE SALIVARY DUCT

Duct cells internalize by wrapping and convergent extension

Duct cells internalize immediately following the secretory cells, beginning with the cells that will form the individual ducts (**Figure 1**)[85]. These cells form tubes by a wrapping type mechanism, whereby the cells become wedge-shaped – wider on the basal side, narrower on the apical side - as they sink below the embryo surface[86]. The internalizing individual duct cells eventually meet and close, beginning distally and extending proximally until they meet up with the common duct primordia at the ventral midline. Once the individual ducts tubes have formed, the common duct tube forms from the more anterior ventral primordia using a similar mechanism. The duct tubes also continue to elongate, eventually forming tubes that are about two cells in diameter for the individual ducts and 3-4 cells for the common duct. Thus, duct elongation occurs by convergent extension[85], a common developmental process that narrows and elongates tissue by cell intercalation[87] and that has been implicated in tube elongation in several Drosophila tissues[88-91] as well as in the vertebrate neural tube[92]. Posterior migration of the gland coupled with the attachment of the duct to the mouthparts may provide the tensile forces required for duct elongation by cell intercalation.

Trh and Eyg are required for duct cell invagination

Trh plays a critical role in internalizing both individual and common ducts since the entire duct fails to invaginate and remains on the ventral surface in *trh* mutant embryos[22,38]. *eye gone* (*eyg*), which encodes a Pax transcription factor positively regulated by Trh, is required to distinguish individual from common duct domains[85]. Eyg is also necessary for morphogenesis of the individual duct tubes[85]; in *eyg* mutants, the individual ducts often fail to elongate to connect the secretory cells to the common duct, resulting in closed internalized secretory tubes that are

disconnected from the rest of the digestive tract. We expect that many more genes that function in duct morphogenesis await discovery.

SALIVARY GLAND FUNCTION: SECRETION AND PRODUCTION OF TISSUE-SPECIFIC GENE PRODUCTS

As the SG undergoes the process of morphogenesis, it simultaneously begins to specialize into a secretory organ. Several transcription factor genes that are induced in the earliest stages of SG formation – *fkh*, *sage and CrebA* - continue to be expressed throughout the life of this organ. These same genes are also expressed in the adult gland, suggesting that they may play similar roles in both the larval and adult tissues. As will be discussed, Fkh, Sage and CrebA play key roles in the main function of the SG – the synthesis and secretion of high levels of protein (**Figure 3**).

Fkh, Sage (and Sens) regulate SG-specific gene products

fkh is required for expression of most SG genes

As mentioned previously, Fkh controls many aspects of SG development, from specification to internalization to gland maintenance. How can Fkh have so many distinct functions in the SG when it is also expressed and required in multiple other tissues[93]? Whereas there may be some commonality in Fkh's role in some tissues, Fkh also has distinct tissue-specific functions. In the SG, tissue-specific Fkh function is largely mediated by Sage[28], a SG-specific bHLH transcription factor.

Sage – a SG-specific bHLH transcription factor – provides specificity to Fkh function

In embryos, Sage (*sa*livary *g*land *E*-box binding protein) is expressed *only* in the SG[94], making it a prime candidate for controlling SG identity and specialization. *sage* null mutant SG cells internalize and form elongated secretory tubes, but undergo massive apoptotic cell death once fully internalized. This finding suggests that even after gland formation and internalization, the

SG must be actively kept alive throughout development. *sage* mutant glands can be rescued from death by co-expression of the anti-apoptotic P35 protein. Although overall gland morphology and polarity is normal, the rescued glands have thin, irregular lumens[28].

Microarray studies reveal that Sage regulates SG-specific secreted or transmembrane proteins and their modifiers. Interestingly, expression of Sage targets entirely depends on Fkh; loss of *fkh* resulted in a complete loss of SG expression of Sage targets. Moreover, experiments expressing either *fkh* or *sage* alone or co-expressing both genes reveal that the combined activity of Fkh and Sage is required for inducing Sage target gene expression in multiple distinct cell types[27,28]. Moreover, Sage, Fkh and another downstream SG transcription factor – Sens – localize to largely overlapping sites on SG polytene chromosomes[28]. Thus, Fkh and Sage (+/- Sens) clearly work together to directly regulate expression of SG-specific genes; the exact mechanism by which these proteins collaborate, however, remains to be elucidated. Similar collaborations between FoxA and Sage-related bHLH proteins are likely to underlie tissue specific gene expression in mammalian dopaminerginic neurons[95,96], pancreas[97-99], as well as the secretory cells of the *C. elegans* pharynx [100].

CrebA upregulates secretory capacity

CrebA and expression of secretory machinery

Early experiments designed to learn how core components of the secretory machinery are regulated in professional secretory cells, such as the SG, revealed that regulation occurs at the transcriptional level. Thirty-four genes encoding protein components of the machinery required at all early steps of the secretory pathway are expressed to significantly higher levels in the SG secretory cells than in surrounding tissues[26]. A combination of in vitro and in vivo DNA binding

experiments, as well as in vivo expression studies, established that CrebA directly regulates secretory pathway component gene (SPCG) expression through a consensus sequence identified by computational analysis of the enhancers for all 34 SPCGs[29]. Genome-wide microarray studies revealed that CrebA is largely dedicated to the regulation of secretory capacity – well over 200 genes encoding core secretory machinery proteins, as well as secreted cargo, require CrebA for their full expression[29]. Moreover, expression of every SPCG that has been tested can be activated in additional cells simply by overexpressing CrebA – or the activated form of any of its five human orthologues, the Creb3/Creb3L family of bZip transcription factors - in those cells[29,101]. Thus, CrebA is both necessary and sufficient for elevated secretory capacity and Creb3/Creb3L have similar activities. CrebA and its human orthologues also appear to boost expression of secreted cargo genes. Whereas direct regulation of cargo genes by the mammalian proteins has been observed[102], studies in flies suggest CrebA may work indirectly by up-regulating expression of Sage (see above); *sage* transcript levels decrease about twofold in *CrebA* mutants[29].

Pasilla encodes a splicing factor also required for high-level SG secretion

Among the early expressed SG genes that come on and stay on in the SG, is *pasilla* (*ps*), which encodes a KH domain-containing nuclear splicing factor related to two mammalian proteins, Nova1 and Nova2[103]. Loss of *ps* in the SG results in late-stage apical lumen irregularities that are linked to a significant reduction in material secreted into the lumen and a corresponding reduction in the size and number of secretory vesicles. Although more than 400 genes have been identified as PS splicing targets in S2 cells, the link to SG secretion remains to be discovered[104].

Diaphanous targets apical secretion

Apical targeting of secretory vesicles in multiple epithelial tubular organs in Drosophila, including the SG, is mediated by Diaphanous (Dia), an actin-nucleation factor that localizes tightly to the apical surface, and by Myosin V (MyoV)[105]. Loss of Dia, which is expressed to high levels in the SG and other secretory organs, or of MyoV has no effect on overall apical-basal polarity but significantly compromises apical secretion. Apical localization of Dia is mediated by its interactions with PIP2 and Rho1, both of which are enriched in the apical membrane[106]. Dia is proposed to nucleate apically directed actin filaments, towards which secretory vesicles are targeted via MyoV-based transport. Mammalian Dia plays a similar role in other secretory organs, including the pancreas and submandibular gland[107].

THE LARVAL-PUPAL SALIVARY GLAND

The larval SG of Drosophila has been used to study several basic cellular functions, including secretion, hormone responsiveness and the cell death pathways. Morphologically, the larval SG looks quite similar to the late embryonic SG, with the salivary duct, imaginal ring cells and secretory gland arranged from proximal to distal (**Figure 6**). In the larva, the proximal region of the secretory gland is termed the transition zone and the cells in the most distal portion of the secretory tube are called the corpus cells. The SG is among the last larval tissues to be destroyed prior to pupation, and its contents are vital to pupae formation as the fly transitions to adulthood.

Ecdysone signaling in the SG

One of the most notable characteristics of the larval SG is its responsiveness to the hormone 20-hydroxy-ecdysone (20E or ecdysone). During the third instar stage, the final stage of larval development, there are three relatively small pulses of ecdysone, followed by a large pulse that signifies the transition from larva to prepupa[108] (**Figure 7**). This final pulse of ecdysone is approximately five times larger than the earlier pulses[109]. An early observation of insect SGs was that the presence of ecdysone led to the formation of puffs along the polytenized chromosomes[110]. Successful in vitro culture of the Drosophila larval SGs allowed for a more complete study of this phenomenon as well as the mapping of the ecdysone-induced puffs to specific chromosomal regions[111,112]. The small pulses of ecdysone induce the intermolt puffs in late larvae that regress during the large ecdysone pulse at the end of the third instar stage[113]. The next set of ecdysone-sensitive puffs has been separated into three distinct groups based on how rapidly they are induced by the large ecdysone pulse: early genes, early-late genes, and late genes. Many early genes encode transcription factors, such as the Broad complex (Br-C), E74,

and E75[114-119]. The early genes, in turn, activate expression of the late genes and repress their own expression[120]. There are many more late genes than early genes, indicating a transcriptional hierarchy in the larval SG not unlike that seen the embryonic gland.

SG glue secretion

The first genes to be induced by the small pulses of ecdysone are those in the intermolt puffs, many of which contain *Salivary glue secretion* (*Sgs*) genes. Expression of at least three *Sgs* genes (Sgs 1, 3, and 4) is controlled by Fkh[128,129] and several *Sgs* genes were identified as potential targets of Sage [130], suggesting that the same transcription factors that cooperate to regulate SG specific expression in embryos also work together to activate larval SG-specific genes. Sgs genes encode glue proteins, several of which are highly glycosylated. Once secreted, glue proteins adhere the pupa to a solid surface during metamorphosis[129,131,132]. The Sgs proteins are synthesized in both the transition and corpus cells of the SG[124], where they accumulate in approximately eleven thousand secretory granules per cell, which coalesce into fewer larger granules prior to their secretion[133]. The final large pulse of ecdysone has two effects on the Sgs genes: transcriptional repression and glue protein secretion[134]. Secretion of the glue granules is dependent on Clathrin, the AP adaptor proteins AP-1, AP47, and EpsinR[135]. Overexpression of E63-1, an early gene that encodes a calcium binding protein, induces premature secretion of glue granules[124]. Moreover, loss of E63-1 and calmodulin impairs secretion[136]. Together, these data suggest a coordinated sequence of events that starts with accumulation of *Sgs* transcripts and glue proteins, followed by a halt in *Sgs* transcription due to increased ecdysone, which also signals the cells to secrete their contents. Without glue protein secretion, the pupa cannot adhere to solid substrates and further development is arrested. In signaling pupa formation, the large ecdysone pulse also signals the larva to destruct (discussed below).

Figure 6. Confocal image of the larval salivary gland with the different cell types artificially colorized. The larval salivary gland includes the large polytenized secretory cells (light blue), the medium sized duct cells (light purple) and the small imaginal ring cells (blue). The fat body (green) attaches to the secretory cells at several places.

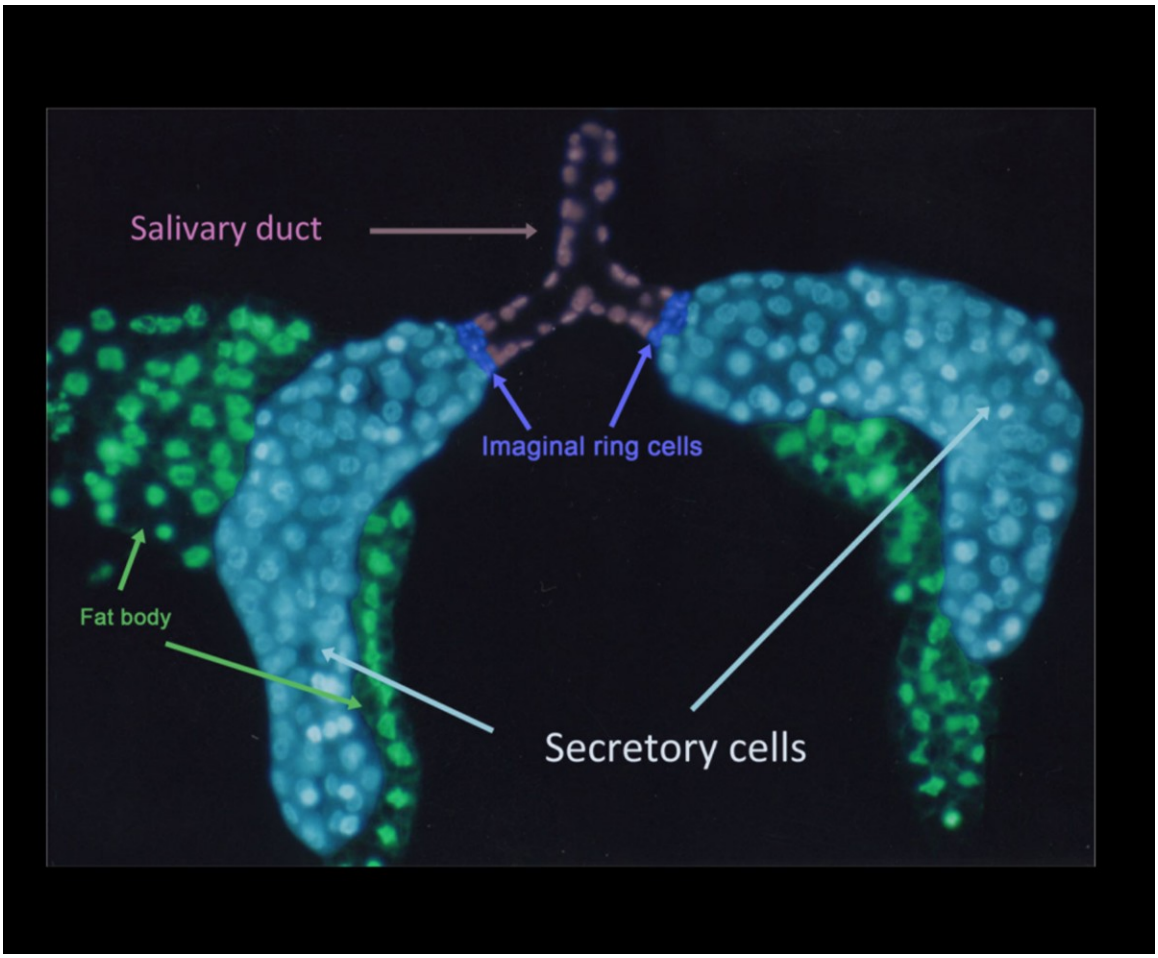

Salivary duct
Imaginal ring cells
Fat body
Secretory cells

Figure 7. Fkh (likely in collaboration with Sage and Sens) keeps the SG alive until the prepupal stage by preventing expression of the apoptosis inducers *reaper* (*rpr*) and *head involution defective* (*hid*). Fkh (and Sage +/- Sens) in combination with low level ecdysone signaling activate transcription of the *Salivary glue secretion* (*Sgs*) genes in late larvae. High-level ecdysone signaling just prior to pupation activates expression of the early ecdysone-responsive genes, which encode transcription factors. A subset of these transcription factors repress *fkh* and *Sgs* transcription and activate expression of genes required for glue secretion. Thus, the glue is secreted when Fkh begins to disappear. In turn, the disappearance of Fkh results in *rpr* and *hid* expression, which overcome DIAP and activate the cell death pathway. Thus, the SG dies shortly after it completes its final task of glue secretion.

Fkh
+ Sage, Sens
rpr, hid
diap1
Sgs*
secretion
20E
20E
20E
20E
Br-C
early genes
E63
(+ targets)
late genes

SG death

In addition to spurring secretion of SG contents, the final large pulse of ecdysone also signals gland death. By this time, many other larval tissues are already partially destroyed. How does the SG remain intact for longer than the surrounding tissues? Two molecules keep the glands alive until after all of their glue is expelled: Fkh, which has kept SG cells alive since the gland was first specified[25], and *D*rosophila *i*nhibitor of *ap*optosis *1* (Diap1)[137]. The late pulse of ecdysone induces Br-C transcription, which, in turn, represses *fkh* transcription and induces the apoptotic cell death cascades. Sustained expression of *fkh* in the SGs delays death, whereas RNAi knockdown of *fkh* at earlier stages induces premature cell death[137]. Although expression of both *rpr* and *hid* RNA are repressed by Fkh, *hid* seems to have the most impact on SG death; overexpression of *hid* alone can induce death and *hid* loss results in gland persistence[138]. Overexpression of other known pro-apoptotic genes - *rpr*, *dronc*, *sickle* or *grim* - does not lead to cell death, at least not on their own, and loss of *rpr* alone does not affect gland persistence[76]. The upregulation of *hid* (and *rpr*) that occurs when Fkh is shut off overcomes the action of Diap1, allowing apoptosis to proceed. Overexpression of p53 alone induces some SG death through apoptotic mechanisms, but is insufficient by itself to cause full gland death[139]. Indeed, SGs also show many signs of autophagic cell death[140]. Overexpression of the autophagy gene *atg1* results in premature cell death[139], which can be rescued by the simultaneous knock down of *atg12*. Moreover, *atg8a* and *atg18* mutant SGS show cleavage of a caspase substrate as well as positive TUNEL staining, suggesting that the apoptotic death pathways remain active when autophagy is blocked. These data reinforce the idea that both autophagy and apoptosis collaborate to ensure the timely death of the SG during metamorphosis.

Genes separate from the canonical death pathways also play important roles in SG death. The matrix metalloprotease Mmp1 is upregulated in the dying glands, whereas its inhibitor Timp (Tissue inhibitor of metalloprotease) is downregulated[141]. Small GTPases are also induced, indicative of cell rearrangements and migration[142]. Mdh2, a mitochondrial malate dehydrogenase, also plays a unique role in cell death[143]. *mdh2* mutant SGs survive longer than wild-type, and although autophagy is initiated in *mdh2* mutant glands, it is not completed. Caspase-3 activation is blocked in *mdh2* mutants, but expression of *rpr*, *hid* and *grim* occurs normally. *mdh2* mutants also have reduced ATP levels relative to wild-type larva. These studies suggest that Mdh2 functions downstream of the previously described death pathways. The findings that Mdh2 and other factors impact SG death suggest that many additional genes function to provide for both the timely death and clean destruction of the larval SG.

CONCLUSIONS

Studies of the Drosophila SG have broader implications regarding how epithelial organs are specified, formed, specialized, maintained and eventually destroyed. Importantly, cell fate is determined by integrating patterning information along both major body axes – anterior-posterior and dorsal-ventral. This is true not only of the SG, but of all other Drosophila tissues that have been studied to the same level of detail, such as the trachea and mesodermal derivatives[144,145]. Although this finding suggests that there may be no true "organ-specifying" genes, clearly there are genes that play major roles in organ development and homeostasis. For the SG, Fkh has this role, affecting all aspects of gland biology. Nonetheless, approximately 40% of SG expressed genes are unaffected by *fkh* loss and Fkh (even with its SG-specific partner Sage) is incapable of stably converting other cell types to a SG fate[11]. Importantly, the only additional cells that persistently express SG markers upon ubiquitous expression of Fkh and Sage, are cells that also express Scr, the Hox gene required for SG formation[28]. To date, Scr itself is the only gene capable of driving ectopic gland formation[1,10].

Excellent progress has been made regarding contributions of the small GTPases and cytoskeletal components to SG morphogenesis, particularly with respect to converting a plate of polarized epithelial cells on the embryo surface into an elongated, fully internalized secretory tube (**Table 1**)[49-51,56,69]. Understanding the initiation and coordination of these events within the gland primordia is the next challenge. We expect that the identification and characterization of early-expressed Fkh targets will be key to fully understanding this process, since *fkh* mutant SGs completely fail to internalize[25].

The SG also provides an excellent model for collective cell migration. The SG is the ultimate collective, since the gland migrates (and elongates) as a fully polarized epithelium. Several pathways and tissues that guide migration have been discovered, but very few of these pathways completely impede migration, leading to the hypothesis that additional guidance molecules exist. Much remains to be learned about how the SG responds to each of the signals and integrates this information to reproducibly arrive at its appropriate final destination, where carrying out its functions is presumably optimized and where the SG in turn can provide cues for positioning other body parts. Understanding the forces fueling tube elongation will also be key.

The past decade has been exciting with regards to learning how SGs specialize – specifically, how SG cells prepare for their major function – high-level secretion – and how SG cells become programmed to produce the right products. High-level secretion is controlled by a single transcription factor – CrebA, which appears to directly activate expression of the *entire* battery of proteins that make up the early secretory machinery[26,29]. Having a single protein (as in flies) or very few proteins (as in humans) with the capacity to coordinately up-regulate the entire secretory pathway provides a simple mechanism for generating sufficient machinery to meet the very different levels of secretory load experienced by various cell types. The beauty of addressing this issue in flies is that with only a single gene with this activity (instead of up to five potentially redundant genes), the consequences of gene loss are much more apparent.

Fkh – like the vertebrate FoxA proteins – is expressed and required in a broad array of embryonic tissues, from cells of the nervous and immune systems to multiple different organ types. Studies of Sage have revealed that Drosophila Fkh achieves SG specificity by partnering up with this tissue-specific bHLH protein[28]. We predict that Fkh will partner with other tissue-

specific proteins (perhaps also other bHLH transcription factors) to regulate completely different sets of targets in the other cells in which it is expressed and required. The next task is to uncover the mechanisms by which Fkh and Sage cooperate to control SG specific gene expression.

Finally, the SG performs vital functions for the animal up to the minutes and hours before its demise. To ensure that these functions can be achieved, Fkh continues to hold death at bay until the final task is completed[137]. Fkh (and likely Sage) work together with the ecdysone-signaling pathway to ensure that not only are the right proteins made at the right time, but that the SG is quickly and cleanly disposed of once its function is accomplished. A relatively complex and seemingly redundant set of events – including what appears to be death by multiple mechanisms – occur to ensure that even death is done right.

ACKNOWLEDGEMENTS

Special thanks to members of the Andrew lab for helpful comments and suggestions.

Our own work in this area has been funded by the RUTH L. KIRSCHSTEIN NATIONAL INDIVIDUAL RESEARCH AWARD 5F31DE022233 and NIH RO1 DE013899.

REFERENCES

1. Panzer S, Weigel D, Beckendorf SK. Organogenesis in Drosophila melanogaster: embryonic salivary gland determination is controlled by homeotic and dorsoventral patterning genes. Development 1992, 114:49-57
2. Henderson KD, Andrew DJ. Regulation and function of Scr, exd, and hth in the Drosophila salivary gland. Developmental biology 2000, 217:362-374
3. Ryoo HD, Mann RS. The control of trunk Hox specificity and activity by Extradenticle. Genes Dev 1999, 13:1704-1716
4. Joshi R, Passner JM, Rohs R, Jain R, Sosinsky A, Crickmore MA, Jacob V, Aggarwal AK, Honig B, Mann RS. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. Cell 2007, 131:530-543
5. Mann RS, Abu-Shaar M. Nuclear import of the homeodomain protein extradenticle in response to Wg and Dpp signalling. Nature 1996, 383:630-633
6. Noro B, Lelli K, Sun L, Mann RS. Competition for cofactor-dependent DNA binding underlies Hox phenotypic suppression. Genes & development 2011, 25:2327-2332
7. Ryoo HD, Marty T, Casares F, Affolter M, Mann RS. Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex. Development 1999, 126:5137-5148
8. Rivas ML, Espinosa-Vazquez JM, Sambrani N, Greig S, Merabet S, Graba Y, Hombria JC. Antagonism versus cooperativity with TALE cofactors at the base of the functional diversification of Hox protein function. PLoS genetics 2013, 9:e1003252
9. Castelli Gair Hombria J, Rivas ML, Sotillos S. Genetic control of morphogenesis - Hox induced organogenesis of the posterior spiracles. The International journal of developmental biology 2009, 53:1349-1358
10. Andrew DJ, Horner MA, Petitt MG, Smolik SM, Scott MP. Setting limits on homeotic gene function: restraint of Sex combs reduced activity by teashirt and other homeotic genes. Embo J 1994, 13:1132-1144
11. Maruyama R, Grevengoed E, Stempniewicz P, Andrew DJ. Genome-wide analysis reveals a major role in cell fate maintenance and an unexpected role in endoreduplication for the Drosophila FoxA gene Fork head. PloS one 2011, 6:e20901
12. Alexandre E, Graba Y, Fasano L, Gallet A, Perrin L, De Zulueta P, Pradel J, Kerridge S, Jacq B. The Drosophila teashirt homeotic protein is a DNA-binding protein and modulo, a HOM-C regulated modifier of variegation, is a likely candidate for being a direct target gene. Mechanisms of development 1996, 59:191-204
13. Fasano L, Roder L, Core N, Alexandre E, Vola C, Jacq B, Kerridge S. The gene teashirt is required for the development of Drosophila embryonic trunk segments and encodes a protein with widely spaced zinc finger motifs. Cell 1991, 64:63-79
14. Roder L, Vola C, Kerridge S. The role of the teashirt gene in trunk segmental identity in Drosophila. Development 1992, 115:1017-1033
15. Taghli-Lamallem O, Gallet A, Leroy F, Malapert P, Vola C, Kerridge S, Fasano L. Direct interaction between Teashirt and Sex combs reduced proteins, via Tsh's acidic domain, is essential for specifying the identity of the prothorax in Drosophila. Developmental biology 2007, 307:142-151
16. Struhl G. Role of the esc+ gene product in ensuring the selective expression of segment-specific homeotic genes in Drosophila. Journal of embryology and experimental morphology 1983, 76:297-331

17. Schneuwly S, Klemenz R, Gehring WJ. Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. Nature 1987, 325:816-818
18. Gibson G, Schier A, LeMotte P, Gehring WJ. The specificities of Sex combs reduced and Antennapedia are defined by a distinct portion of each protein that includes the homeodomain. Cell 1990, 62:1087-1103
19. Gonzalez-Reyes A, Morata G. Organization of the Drosophila head as revealed by the ectopic expression of the Ultrabithorax product. Development 1991, 113:1459-1471
20. Gonzalez-Reyes A, Urquia N, Gehring WJ, Struhl G, Morata G. Are cross-regulatory interactions between homoeotic genes functionally significant? Nature 1990, 344:78-80
21. Mann RS, Hogness DS. Functional dissection of Ultrabithorax proteins in D. melanogaster. Cell 1990, 60:597-610
22. Isaac DD, Andrew DJ. Tubulogenesis in Drosophila: a requirement for the trachealess gene product. Genes & development 1996, 10:103-117
23. Henderson KD, Isaac DD, Andrew DJ. Cell fate specification in the Drosophila salivary gland: the integration of homeotic gene function with the DPP signaling cascade. Developmental biology 1999, 205:10-21
24. Zhou B, Bagri A, Beckendorf SK. Salivary gland determination in Drosophila: a salivary-specific, fork head enhancer integrates spatial pattern and allows fork head autoregulation. Developmental biology 2001, 237:54-67
25. Myat MM, Andrew DJ. Fork head prevents apoptosis and promotes cell shape change during formation of the Drosophila salivary glands. Development 2000, 127:4217-4226
26. Abrams EW, Andrew DJ. CrebA regulates secretory activity in the Drosophila salivary gland and epidermis. Development 2005, 132:2743-2758
27. Abrams EW, Mihoulides WK, Andrew DJ. Fork head and Sage maintain a uniform and patent salivary gland lumen through regulation of two downstream target genes, PH4alphaSG1 and PH4alphaSG2. Development 2006, 133:3517-3527
28. Fox RM, Vaishnavi A, Maruyama R, Andrew DJ. Organ-specific gene expression: the bHLH protein Sage provides tissue specificity to Drosophila FoxA. Development 2013, 140:2160-2171
29. Fox RM, Hanlon CD, Andrew DJ. The CrebA/Creb3-like transcription factors are major and direct regulators of secretory capacity. J Cell Biol 2010, 191:479-492
30. Chandrasekaran V, Beckendorf SK. senseless is necessary for the survival of embryonic salivary glands in Drosophila. Development 2003, 130:4719-4728
31. Fleming RJ, Scottgale TN, Diederich RJ, Artavanis-Tsakonas S. The gene Serrate encodes a putative EGF-like transmembrane protein essential for proper ectodermal development in Drosophila melanogaster. Genes & development 1990, 4:2188-2201
32. Thomas U, Speicher SA, Knust E. The Drosophila gene Serrate encodes an EGF-like transmembrane protein with a complex expression pattern in embryos and wing discs. Development 1991, 111:749-761
33. Klambt C, Glazer L, Shilo BZ. breathless, a Drosophila FGF receptor homolog, is essential for migration of tracheal and specific midline glial cells. Genes & development 1992, 6:1668-1678
34. Gregory SL, Kortschak RD, Kalionis B, Saint R. Characterization of the dead ringer gene identifies a novel, highly conserved family of sequence-specific DNA-binding proteins. Molecular and cellular biology 1996, 16:792-799
35. Kasai Y, Stahl S, Crews S. Specification of the Drosophila CNS midline cell lineage: direct control of single-minded transcription by dorsal/ventral patterning genes. Gene expression 1998, 7:171-189

36. Nambu JR, Franks RG, Hu S, Crews ST. The single-minded gene of Drosophila is required for the expression of genes important for the development of CNS midline cells. Cell 1990, 63:63-75
37. Klambt C. EGF receptor signalling: roles of star and rhomboid revealed. Current biology : CB 2002, 12:R21-23
38. Kuo YM, Jones N, Zhou B, Panzer S, Larson V, Beckendorf SK. Salivary duct determination in Drosophila: roles of the EGF receptor signalling pathway and the transcription factors fork head and trachealess. Development 1996, 122:1909-1917
39. Haberman AS, Isaac DD, Andrew DJ. Specification of cell fates within the salivary gland primordium. Dev Biol 2003, 258:443-453
40. Kidd S, Baylies MK, Gasic GP, Young MW. Structure and distribution of the Notch protein in developing Drosophila. Genes & development 1989, 3:1113-1129
41. Yamamoto S, Charng WL, Rana NA, Kakuda S, Jaiswal M, Bayat V, Xiong B, Zhang K, Sandoval H, David G, Wang H, Haltiwanger RS, Bellen HJ. A mutation in EGF repeat-8 of Notch discriminates between Serrate/Jagged and Delta family ligands. Science 2012, 338:1229-1232
42. Gridley T. Notch signaling in the vasculature. Current topics in developmental biology 2010, 92:277-309
43. Nantie LB, Himes AD, Getz DR, Raetzman LT. Notch signaling in postnatal pituitary expansion: proliferation, progenitors and cell specification. Molecular endocrinology 2014, me20131425
44. Murtaugh LC, Stanger BZ, Kwan KM, Melton DA. Notch signaling controls multiple steps of pancreatic differentiation. Proceedings of the National Academy of Sciences of the United States of America 2003, 100:14920-14925
45. Patten BA, Peyrin JM, Weinmaster G, Corfas G. Sequential signaling through Notch1 and erbB receptors mediates radial glia differentiation. The Journal of neuroscience : the official journal of the Society for Neuroscience 2003, 23:6132-6140
46. Myat MM, Andrew DJ. Organ shape in the Drosophila salivary gland is controlled by regulated, sequential internalization of the primordia. Development 2000, 127:679-691
47. Manning AJ, Peters KA, Peifer M, Rogers SL. Regulation of epithelial morphogenesis by the G protein-coupled receptor mist and its ligand fog. Science signaling 2013, 6:ra98
48. Rogers SL, Wiedemann U, Hacker U, Turck C, Vale RD. Drosophila RhoGEF2 associates with microtubule plus ends in an EB1-dependent manner. Current biology : CB 2004, 14:1827-1833
49. Nikolaidou KK, Barrett K. A Rho GTPase signaling pathway is used reiteratively in epithelial folding and potentially selects the outcome of Rho activation. Current biology : CB 2004, 14:1822-1826
50. Kolesnikov T, Beckendorf SK. 18 wheeler regulates apical constriction of salivary gland cells via the Rho-GTPase-signaling pathway. Dev Biol 2007, 307:53-61
51. Xu N, Keung B, Myat MM. Rho GTPase controls invagination and cohesive migration of the Drosophila salivary gland through Crumbs and Rho-kinase. Developmental biology 2008, 321:88-100
52. Tepass U, Theres C, Knust E. crumbs encodes an EGF-like protein expressed on apical membranes of Drosophila epithelial cells and required for organization of epithelia. Cell 1990, 61:787-799
53. Tepass U, Knust E. Crumbs and stardust act in a genetic pathway that controls the organization of epithelia in Drosophila melanogaster. Developmental biology 1993, 159:311-326

54. Wodarz A, Hinz U, Engelbert M, Knust E. Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of Drosophila. Cell 1995, 82:67-76
55. Tepass U. Crumbs, a component of the apical membrane, is required for zonula adherens formation in primary epithelia of Drosophila. Dev Biol 1996, 177:217-225
56. Roper K. Anisotropy of Crumbs and aPKC drives myosin cable assembly during tube formation. Developmental cell 2012, 23:939-953
57. Chandrasekaran V, Beckendorf SK. Tec29 controls actin remodeling and endoreplication during invagination of the Drosophila embryonic salivary glands. Development 2005, 132:3515-3524
58. Patel U, Myat MM. Receptor guanylyl cyclase Gyc76C is required for invagination, collective migration and lumen shape in the Drosophila embryonic salivary gland. Biology open 2013, 2:711-717
59. Myat MM, Andrew DJ. Epithelial tube morphology is determined by the polarized growth and delivery of apical membrane. Cell 2002, 111:879-891
60. Bronner G, Jackle H. Regulation and function of the terminal gap gene huckebein in the Drosophila blastoderm. The International journal of developmental biology 1996, 40:157-165
61. Mosley-Bishop KL, Li Q, Patterson L, Fischer JA. Molecular analysis of the klarsicht gene and its role in nuclear migration within differentiating cells of the Drosophila eye. Current biology : CB 1999, 9:1211-1220
62. Welte MA, Gross SP, Postner M, Block SM, Wieschaus EF. Developmental regulation of vesicle transport in Drosophila embryos: forces and kinetics. Cell 1998, 92:547-557
63. Bradley PL, Andrew DJ. ribbon encodes a novel BTB/POZ protein required for directed cell migration in Drosophila melanogaster. Development 2001, 128:3001-3015
64. Shim K, Blake KJ, Jack J, Krasnow MA. The Drosophila ribbon gene encodes a nuclear BTB domain protein that promotes epithelial migration and morphogenesis. Development 2001, 128:4923-4933
65. Kerman BE, Cheshire AM, Myat MM, Andrew DJ. Ribbon modulates apical membrane during tube elongation through Crumbs and Moesin. Developmental biology 2008, 320:278-288
66. Cheshire AM, Kerman BE, Zipfel WR, Spector AA, Andrew DJ. Kinetic and mechanical analysis of live tube morphogenesis. Developmental dynamics : an official publication of the American Association of Anatomists 2008, 237:2874-2888
67. Medina E, Williams J, Klipfell E, Zarnescu D, Thomas G, Le Bivic A. Crumbs interacts with moesin and beta(Heavy)-spectrin in the apical membrane skeleton of Drosophila. J Cell Biol 2002, 158:941-951
68. Polesello C, Delon I, Valenti P, Ferrer P, Payre F. Dmoesin controls actin-based cell shape and polarity during Drosophila melanogaster oogenesis. Nature cell biology 2002, 4:782-789
69. Xu N, Bagumian G, Galiano M, Myat MM. Rho GTPase controls Drosophila salivary gland lumen size through regulation of the actin cytoskeleton and Moesin. Development 2011, 138:5415-5427
70. Reed RA, Womble MA, Dush MK, Tull RR, Bloom SK, Morckel AR, Devlin EW, Nascone-Yoder NM. Morphogenesis of the primitive gut tube is generated by Rho/ROCK/myosin II-mediated endoderm rearrangements. Developmental dynamics : an official publication of the American Association of Anatomists 2009, 238:3111-3125

71. Tepass U, Hartenstein V. The development of cellular junctions in the Drosophila embryo. Developmental biology 1994, 161:563-596
72. Pirraglia C, Jattani R, Myat MM. Rac function in epithelial tube morphogenesis. Developmental biology 2006, 290:435-446
73. Bokoch GM. Biology of the p21-activated kinases. Annual review of biochemistry 2003, 72:743-781
74. Arias-Romero LE, Chernoff J. A tale of two Paks. Biology of the cell / under the auspices of the European Cell Biology Organization 2008, 100:97-108
75. Pirraglia C, Walters J, Myat MM. Pak1 control of E-cadherin endocytosis regulates salivary gland lumen size and shape. Development 2010, 137:4177-4189
76. Yin VP, Thummel CS. Mechanisms of steroid-triggered programmed cell death in Drosophila. Seminars in cell & developmental biology 2005, 16:237-243
77. Ismat A, Cheshire AM, Andrew DJ. The secreted AdamTS-A metalloprotease is required for collective cell migration. Development 2013, 140:1981-1993
78. Schlichting K, Wilsch-Brauninger M, Demontis F, Dahmann C. Cadherin Cad99C is required for normal microvilli morphology in Drosophila follicle cells. Journal of cell science 2006, 119:1184-1195
79. D'Alterio C, Tran DD, Yeung MW, Hwang MS, Li MA, Arana CJ, Mulligan VK, Kubesh M, Sharma P, Chase M, Tepass U, Godt D. Drosophila melanogaster Cad99C, the orthologue of human Usher cadherin PCDH15, regulates the length of microvilli. J Cell Biol 2005, 171:549-558
80. Chung SY, Andrew DJ. Cadherin 99C regulates apical expansion and cell rearrangement during epithelial tube elongation. Development 2014,
81. Bradley PL, Myat MM, Comeaux CA, Andrew DJ. Posterior migration of the salivary gland requires an intact visceral mesoderm and integrin function. Dev Biol 2003, 257:249-262
82. Vining MS, Bradley PL, Comeaux CA, Andrew DJ. Organ positioning in Drosophila requires complex tissue-tissue interactions. Dev Biol 2005, 287:19-34
83. Harris KE, Schnittke N, Beckendorf SK. Two ligands signal through the Drosophila PDGF/VEGF receptor to ensure proper salivary gland positioning. Mechanisms of development 2007, 124:441-448
84. Kolesnikov T, Beckendorf SK. NETRIN and SLIT guide salivary gland migration. Dev Biol 2005, 284:102-111
85. Jones NA, Kuo YM, Sun YH, Beckendorf SK. The Drosophila Pax gene eye gone is required for embryonic salivary duct development. Development 1998, 125:4163-4174
86. Kerman BE, Andrew DJ. Staying alive: dalmation mediated blocking of apoptosis is essential for tissue maintenance. Developmental dynamics : an official publication of the American Association of Anatomists 2010, 239:1609-1621
87. Irvine KD, Wieschaus E. Cell intercalation during Drosophila germband extension and its regulation by pair-rule segmentation genes. Development 1994, 120:827-841
88. Iwaki DD, Johansen KA, Singer JB, Lengyel JA. drumstick, bowl, and lines are required for patterning and cell rearrangement in the Drosophila embryonic hindgut. Developmental biology 2001, 240:611-626
89. Lengyel JA, Iwaki DD. It takes guts: the Drosophila hindgut as a model system for organogenesis. Developmental biology 2002, 243:1-19
90. Jung AC, Denholm B, Skaer H, Affolter M. Renal tubule development in Drosophila: a closer look at the cellular level. Journal of the American Society of Nephrology : JASN 2005, 16:322-328

91. Caussinus E, Colombelli J, Affolter M. Tip-cell migration controls stalk-cell intercalation during Drosophila tracheal tube elongation. Current biology : CB 2008, 18:1727-1734
92. Keller R, Davidson L, Edlund A, Elul T, Ezin M, Shook D, Skoglund P. Mechanisms of convergence and extension by cell intercalation. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 2000, 355:897-922
93. Weigel D, Jurgens G, Kuttner F, Seifert E, Jackle H. The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. Cell 1989, 57:645-658
94. Moore AW, Barbel S, Jan LY, Jan YN. A genomewide survey of basic helix-loop-helix factors in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 2000, 97:10436-10441
95. Kele J, Simplicio N, Ferri AL, Mira H, Guillemot F, Arenas E, Ang SL. Neurogenin 2 is required for the development of ventral midbrain dopaminergic neurons. Development 2006, 133:495-505
96. Lin W, Metzakopian E, Mavromatakis YE, Gao N, Balaskas N, Sasaki H, Briscoe J, Whitsett JA, Goulding M, Kaestner KH, Ang SL. Foxa1 and Foxa2 function both upstream of and cooperatively with Lmx1a and Lmx1b in a feedforward loop promoting mesodiencephalic dopaminergic neuron development. Developmental biology 2009, 333:386-396
97. Krapp A, Knofler M, Frutiger S, Hughes GJ, Hagenbuchle O, Wellauer PK. The p48 DNA-binding subunit of transcription factor PTF1 is a new exocrine pancreas-specific basic helix-loop-helix protein. The EMBO journal 1996, 15:4317-4329
98. Krapp A, Knofler M, Ledermann B, Burki K, Berney C, Zoerkler N, Hagenbuchle O, Wellauer PK. The bHLH protein PTF1-p48 is essential for the formation of the exocrine and the correct spatial organization of the endocrine pancreas. Genes & development 1998, 12:3752-3763
99. Gao N, Le Lay J, Qin W, Doliba N, Schug J, Fox AJ, Smirnova O, Matschinsky FM, Kaestner KH. Foxa1 and Foxa2 maintain the metabolic and secretory features of the mature beta-cell. Molecular endocrinology 2010, 24:1594-1604
100. Smit RB, Schnabel R, Gaudet J. The HLH-6 transcription factor regulates C. elegans pharyngeal gland development and function. PLoS genetics 2008, 4:e1000222
101. Barbosa S, Fasanella G, Carreira S, Llarena M, Fox R, Barreca C, Andrew D, O'Hare P. An orchestrated program regulating secretory pathway genes and cargos by the transmembrane transcription factor CREB-H. Traffic 2013, 14:382-398
102. Murakami T, Saito A, Hino S, Kondo S, Kanemoto S, Chihara K, Sekiya H, Tsumagari K, Ochiai K, Yoshinaga K, Saitoh M, Nishimura R, Yoneda T, Kou I, Furuichi T, Ikegawa S, Ikawa M, Okabe M, Wanaka A, Imaizumi K. Signalling mediated by the endoplasmic reticulum stress transducer OASIS is involved in bone formation. Nature cell biology 2009, 11:1205-1211
103. Seshaiah P, Miller B, Myat MM, Andrew DJ. pasilla, the Drosophila homologue of the human Nova-1 and Nova-2 proteins, is required for normal secretion in the salivary gland. Developmental biology 2001, 239:309-322
104. Brooks AN, Yang L, Duff MO, Hansen KD, Park JW, Dudoit S, Brenner SE, Graveley BR. Conservation of an RNA regulatory map between Drosophila and mammals. Genome research 2011, 21:193-202
105. Massarwa R, Schejter ED, Shilo BZ. Apical secretion in epithelial tubes of the Drosophila embryo is directed by the Formin-family protein Diaphanous. Developmental cell 2009, 16:877-888

106. Rousso T, Shewan AM, Mostov KE, Schejter ED, Shilo BZ. Apical targeting of the formin Diaphanous in Drosophila tubular epithelia. eLife 2013, 2:e00666
107. Geron E, Schejter ED, Shilo BZ. Directing exocrine secretory vesicles to the apical membrane by actin cables generated by the formin mDia1. Proceedings of the National Academy of Sciences of the United States of America 2013, 110:10652-10657
108. Redfern CPF. Ecdysteroid synthesis by the ring gland of Drosophila melanogaster during late-larval, prepupal and pupal development. Journal of insect physiology 1983, 29:65-71
109. Warren JT, Yerushalmi Y, Shimell MJ, O'Connor MB, Restifo LL, Gilbert LI. Discrete pulses of molting hormone, 20-hydroxyecdysone, during late larval development of Drosophila melanogaster: correlations with changes in gene activity. Developmental dynamics : an official publication of the American Association of Anatomists 2006, 235:315-326
110. Clever U, Karlson P. [Induction of puff changes in the salivary gland chromosomes of Chironomus tentans by ecdysone]. Experimental cell research 1960, 20:623-626
111. Ashburner M, Lemeunier F. Patterns of puffing activity in the salivary gland chromosomes of Drosophila. VII. Homology of puffing patterns on chromosome arm 3L in D. melanogaster and D. yakuba, with notes on puffing in D. teissieri. Chromosoma 1972, 38:283-295
112. Ashburner M. Patterns of puffing activity in the salivary gland chromosomes of Drosophila. VI. Induction by ecdysone in salivary glands of D. melanogaster cultured in vitro. Chromosoma 1972, 38:255-281
113. Lehmann M, Korge G. Ecdysone regulation of the Drosophila Sgs-4 gene is mediated by the synergistic action of ecdysone receptor and SEBP 3. Embo J 1995, 14:716-726
114. Crossgrove K, Bayer CA, Fristrom JW, Guild GM. The Drosophila Broad-Complex early gene directly regulates late gene transcription during the ecdysone-induced puffing cascade. Developmental biology 1996, 180:745-758
115. Fletcher JC, D'Avino PP, Thummel CS. A steroid-triggered switch in E74 transcription factor isoforms regulates the timing of secondary-response gene expression. Proceedings of the National Academy of Sciences of the United States of America 1997, 94:4582-4586
116. Sun G, Zhu J, Li C, Tu Z, Raikhel AS. Two isoforms of the early E74 gene, an Ets transcription factor homologue, are implicated in the ecdysteroid hierarchy governing vitellogenesis of the mosquito, Aedes aegypti. Molecular and cellular endocrinology 2002, 190:147-157
117. Zhou B, Hiruma K, Jindra M, Shinoda T, Segraves WA, Malone F, Riddiford LM. Regulation of the transcription factor E75 by 20-hydroxyecdysone and juvenile hormone in the epidermis of the tobacco hornworm, Manduca sexta, during larval molting and metamorphosis. Developmental biology 1998, 193:127-138
118. Segraves WA, Hogness DS. The E75 ecdysone-inducible gene responsible for the 75B early puff in Drosophila encodes two new members of the steroid receptor superfamily. Genes & development 1990, 4:204-219
119. Burtis KC, Thummel CS, Jones CW, Karim FD, Hogness DS. The Drosophila 74EF early puff contains E74, a complex ecdysone-inducible gene that encodes two ets-related proteins. Cell 1990, 61:85-99
120. Ashburner M. Sequential gene activation by ecdysone in polytene chromosomes of Drosophila melanogaster. II. The effects of inhibitors of protein synthesis. Developmental biology 1974, 39:141-157

121. Koelle MR, Talbot WS, Segraves WA, Bender MT, Cherbas P, Hogness DS. The Drosophila EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. Cell 1991, 67:59-77
122. King-Jones K, Charles JP, Lam G, Thummel CS. The ecdysone-induced DHR4 orphan nuclear receptor coordinates growth and maturation in Drosophila. Cell 2005, 121:773-784
123. Talbot WS, Swyryd EA, Hogness DS. Drosophila tissues with different metamorphic responses to ecdysone express different ecdysone receptor isoforms. Cell 1993, 73:1323-1337
124. Biyasheva A, Do TV, Lu Y, Vaskova M, Andres AJ. Glue secretion in the Drosophila salivary gland: a model for steroid-regulated exocytosis. Developmental biology 2001, 231:234-251
125. Yao TP, Forman BM, Jiang Z, Cherbas L, Chen JD, McKeown M, Cherbas P, Evans RM. Functional ecdysone receptor is the product of EcR and Ultraspiracle genes. Nature 1993, 366:476-479
126. Costantino BF, Bricker DK, Alexandre K, Shen K, Merriam JR, Antoniewski C, Callender JL, Henrich VC, Presente A, Andres AJ. A novel ecdysone receptor mediates steroid-regulated developmental events during the mid-third instar of Drosophila. PLoS genetics 2008, 4:e1000102
127. Costantino BF. The Larval salivary gland of Drosophila melanogaster: A model system for temporal and spatial steroid hormone regulation. UNLV Theses/Dissertations/Professional Papers/Capstones 2008,
128. Mach V, Ohno K, Kokubo H, Suzuki Y. The Drosophila fork head factor directly controls larval salivary gland-specific expression of the glue protein gene Sgs3. Nucleic acids research 1996, 24:2387-2394
129. Roth GE, Wattler S, Bornschein H, Lehmann M, Korge G. Structure and regulation of the salivary gland secretion protein gene Sgs-1 of Drosophila melanogaster. Genetics 1999, 153:753-762
130. Li TR, White KP. Tissue-specific gene expression and ecdysone-regulated genomic networks in Drosophila. Developmental cell 2003, 5:59-72
131. Muskavitch MA, Hogness DS. An expandable gene that encodes a Drosophila glue protein is not expressed in variants lacking remote upstream sequences. Cell 1982, 29:1041-1051
132. Garfinkel MD, Pruitt RE, Meyerowitz EM. DNA sequences, gene regulation and modular protein evolution in the Drosophila 68C glue gene cluster. Journal of molecular biology 1983, 168:765-789
133. Farkas R, Suakova G. Developmental regulation of granule size and numbers in larval salivary glands of drosophila by steroid hormone ecdysone. Cell biology international 1999, 23:671-676
134. Andres AJ, Fletcher JC, Karim FD, Thummel CS. Molecular analysis of the initiation of insect metamorphosis: a comparative study of Drosophila ecdysteroid-regulated transcription. Developmental biology 1993, 160:388-404
135. Burgess J, Jauregui M, Tan J, Rollins J, Lallet S, Leventis PA, Boulianne GL, Chang HC, Le Borgne R, Kramer H, Brill JA. AP-1 and clathrin are essential for secretory granule biogenesis in Drosophila. Molecular biology of the cell 2011, 22:2094-2105
136. Thummel CS. Ecdysone-regulated puff genes 2000. Insect biochemistry and molecular biology 2002, 32:113-120

137. Cao C, Liu Y, Lehmann M. Fork head controls the timing and tissue selectivity of steroid-induced developmental cell death. J Cell Biol 2007, 176:843-852
138. Juhasz G, Sass M. Hid can induce, but is not required for autophagy in polyploid larval Drosophila tissues. European journal of cell biology 2005, 84:491-502
139. Berry DL, Baehrecke EH. Growth arrest and autophagy are required for salivary gland cell degradation in Drosophila. Cell 2007, 131:1137-1148
140. Lee CY, Simon CR, Woodard CT, Baehrecke EH. Genetic mechanism for the stage- and tissue-specific regulation of steroid triggered programmed cell death in Drosophila. Developmental biology 2002, 252:138-148
141. Gorski SM, Chittaranjan S, Pleasance ED, Freeman JD, Anderson CL, Varhol RJ, Coughlin SM, Zuyderduyn SD, Jones SJ, Marra MA. A SAGE approach to discovery of genes involved in autophagic cell death. Current biology : CB 2003, 13:358-363
142. Lee CY, Clough EA, Yellon P, Teslovich TM, Stephan DA, Baehrecke EH. Genome-wide analyses of steroid- and radiation-triggered programmed cell death in Drosophila. Current biology : CB 2003, 13:350-357
143. Wang L, Lam G, Thummel CS. Med24 and Mdh2 are required for Drosophila larval salivary gland cell death. Developmental dynamics : an official publication of the American Association of Anatomists 2010, 239:954-964
144. Lee HH, Frasch M. Nuclear integration of positive Dpp signals, antagonistic Wg inputs and mesodermal competence factors during Drosophila visceral mesoderm induction. Development 2005, 132:1429-1442
145. Maruyama R, Andrew DJ. Drosophila as a model for epithelial tube formation. Developmental dynamics : an official publication of the American Association of Anatomists 2012, 241:119-135

CHAPTER 2:

Outside – In Signaling: A Review of the *Drosophila* GPCR family

ABSTRACT

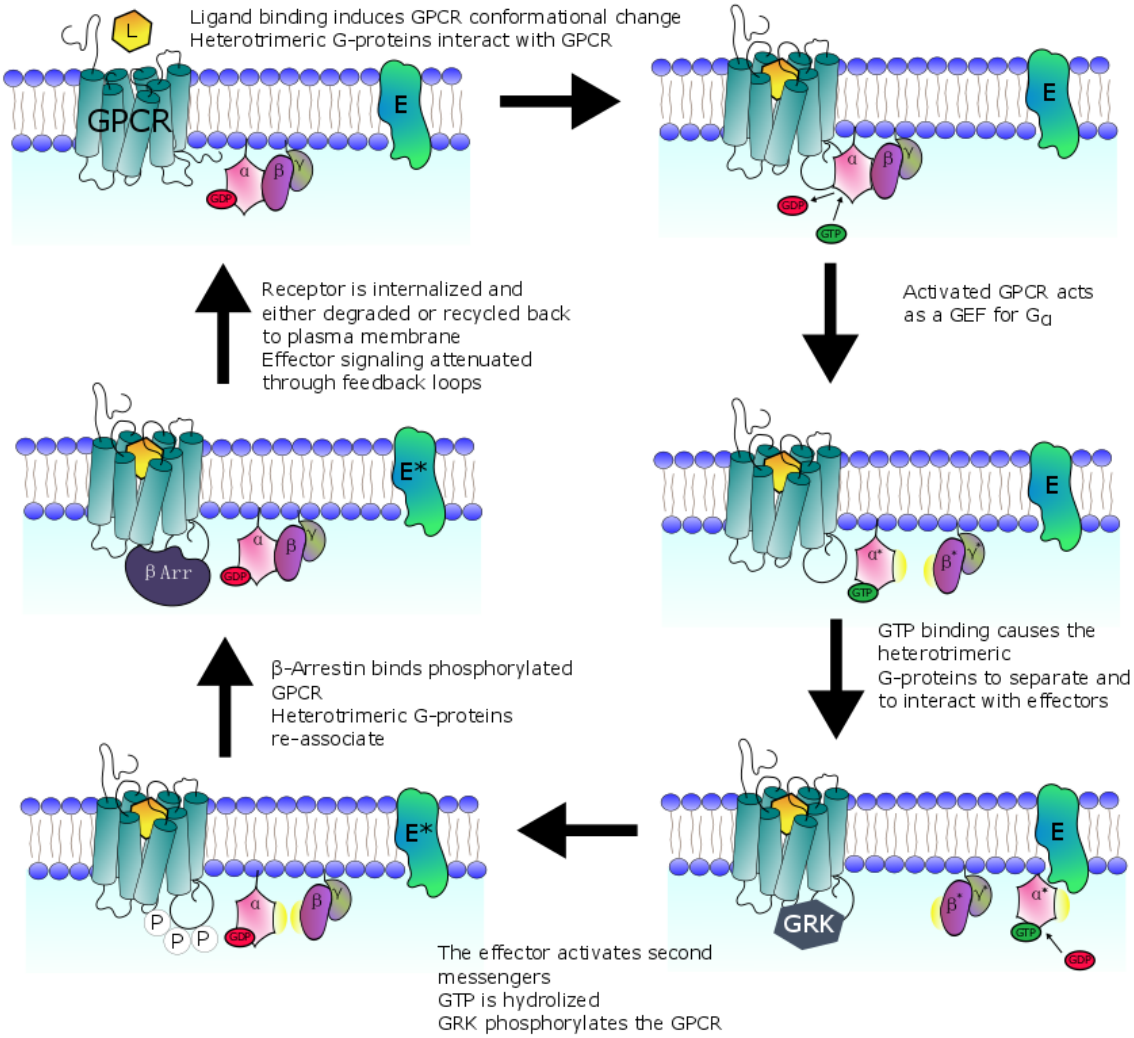
G-protein coupled receptors (GPCRs) are the largest family of receptors in many organisms, including worms, mice, and humans. GPCRs are seven-transmembrane spanning proteins that are activated by binding a stimulus (or ligand) in the extracellular space. They transduce information to the inside of the cell through conformational changes. In turn, the conformational changes activate heterotrimeric G-proteins, which execute the downstream signaling pathways. The ligand-GPCR interaction is highly specific, and an efficient cellular response is vital for the health of the cell and organism. In this Commentary, we focus on the *Drosophila* GPCRs, which are not as well-characterized as their worm and mammalian counterparts. We also present here a phylogenetic analysis which has led us to uncover several interesting relationships among *Drosophila* GPCRs, such as the clustering of several taste receptors with neurotransmitter receptors. Because many GPCRs remain uncharacterized orphan receptors, this analysis suggests potential roles for several family members. Finally, we discuss recently emerged roles of GPCRs in *Drosophila* embryogenesis, a field we expect will uncover many previously unappreciated functions of GPCRs.

Introduction to G-protein coupled receptors – finely tuned environmental sensors

How cells sense and respond to outside stimuli has been a key question in biology for well over a century. The hypothesized existence of receptors that span the cell membrane and are able to both sense and transduce signals seems logical today, but when this idea was first proposed, it was summarily rejected (Lefkowitz, 2013). Decades of biochemical studies have since proven that such receptors exist, and that most are G-protein coupled receptors (GPCRs). The main feature of GPCRs is their seven transmembrane-spanning segments, which position the N-terminus of the protein on the outside of the cell and the C-terminus inside. GPCRs bind an astoundingly diverse set of ligands – proteins, peptides, hormones, drugs, photons – usually by capturing the ligand with their N-terminus and/or within a pocket formed by the extracellular and transmembrane domains (**Fig. 1**). The GPCR cycle, described in more detail below, is an elegant cellular solution for sensing a specific exogenous signal, transducing it to a signaling cascade, and then terminating the signal.

GPCRs are widely represented in all forms of life, from bacteria to plants and animals, including all of the major model organisms. Over 800 GPCRs are encoded in the human genome and well over 700 in the zebrafish (Fredrikkson and Schioth, 2005). *Caernorhabditis elegans* is predicted to encode over 1000 GPCRs, a particularly impressive number, as this figure accounts for over five percent of the entire worm genome. Mice also encode a large number of GPCRs (over 1300). *Drosophila* encodes over 200 GPCRs, and just over 50 are found in *Dictyostelium* (Prabhu and Eichinger, 2006). In contrast, yeast encode a surprisingly small number of GPCRs – three in *Saccharomyces cerevisciae* and nine in *S. pombe*. One kingdom where GPCRs were thought to be conspicuously absent was the plants, due to lack of encoded G-protein exchange factors (GEFs) (Urano et al 2012); more recent work has, however, identified at least twelve plant GPCRs (Taddese et al 2014), restoring GPCRs' status as ubiquitous signaling systems.

Figure1: The G-protein coupled receptor (GPCR) cycle. The GPCR cycle starts on the top left of the figure. In its basal state, a GPCR is free of ligand. *Gα* binds to GDP and associates with Gβγ. The heterotrimeric protein complex may associate with the receptor at this point, or remain free in the membrane as pictured. Upon ligand binding the GPCR becomes activated and undergoes a conformational change. The activated GPCR acts as a GEF for *Gα*. The resulting GTP-bound *Gα* separates from βγ, and the heterotrimeric proteins are active (*). Activated *Gα* can then interact with an effector (E), such as phospholipase C) (PLC) or adenylate cyclase, which results in effector activation (*) and initiation of a second- messenger cascade. The GTP in *Gα* is then hydrolyzed to GDP through the activity of Gα and RGS proteins (not shown), leading to *Gα* inactivation and reassociation of heterotrimeric protein complex. Independently, GRKs bind to and phosphorylate the GPCR. This stimulates its binding by β-arrestin, which promotes internalization of the receptor. The GPCR can then be recycled back to the cell surface without ligand, restarting the cycle.


Ligand binding induces GPCR conformational change
Heterotrimeric G-proteins interact with GPCR
L
GPCR
E
α
β
γ
GDP
GTP
Activated GPCR acts as a GEF for Gα
α*
β*
γ*
GTP binding causes the heterotrimeric G-proteins to separate and to interact with effectors
GRK
The effector activates second messengers
GTP is hydrolized
GRK phosphorylates the GPCR
E*
P
β-Arrestin binds phosphorylated GPCR
Heterotrimeric G-proteins re-associate
β Arr
Receptor is internalized and either degraded or recycled back to plasma membrane
Effector signaling attenuated through feedback loops

GPCRs are involved in nearly every aspect of life, from early development to heart function to neuronal activity (Wettschurek et al 2005). Mutations in GPCRs are linked to a number of human diseases, such as Usher syndrome, which results in variable onset deaf-blindness (Schoneberg et al 2004). Cell migration is another process that requires GPCRs, in both beneficial and detrimental ways. The single-celled amoeba Dictyostelium uses four GPCRs, cAR1-4, to detect cAMP, triggering migration and coalescence into a multicellular organism (Prabhu and Eichinger, 2006). In both zebrafish and mice, germ cell migration is regulated by the chemokine receptor CXCR4 and its ligand stromal derived factor 1 (SDF1) (Doitsidou et al., 2002; Molyneaux et al 2003). Neutrophils also migrate via activation of a GPCR (Becker et al 1987). GPCR-regulated cell migration can also be detrimental for the organism, primarily during cancer (Lappano et al 2011). Overly activated GPCRs are able to trans-activate EGFR and other receptors, which can cause unregulated growth and cell migration.

In this Commentary, we first cover the basics of GPCR signaling, from how the receptor is arranged in the membrane to heterotrimeric protein activation and receptor recycling. We focus then on the *Drosophila* GPCR family, including the gustatory and odorant receptors. Phylogenetic trees representing the entire repertoire of GPCRs in *Drosophila* are presented, revealing interesting relationships among GPCRs previously considered to be distantly related. Moreover, this analysis suggests potential roles for some of the many uncharacterized GPCRs. Finally, we discuss the known roles of GPCRs in *Drosophila* development, an open and exciting field given the number of GPCRs encoded in the fly genome.

GPCR Structure-Function

Because the GPCR superfamily is so diverse, there is little sequence conservation among families. Nonetheless, the superfamily does share several architectural features. The N-terminus and extracellular loops (ECL) are responsible for ligand binding. This can involve direct binding of the ECL to the ligand (as is the case for metabotropic Glutamate (mGlu) receptors) or funneling of a hydrophobic ligand into a binding pocket formed by the transmembrane domains (Venkatakrishnan et al 2013). ECLs often contain disulphide bridges to stabilize the loops and prevent promiscuous GPCR signaling. Indeed, ligand-binding does not induce a simple on-off state for GPCRs. GPCRs are dynamic proteins that fluctuate between many states (**Fig. 1**). Ligand binding stabilizes the GPCR into an "on" position, which is further stabilized by binding of a G-protein (Kobilka, 2012).

Bioinformatic analyses show that residues, such as Asparagine, Tryptophan, and Proline, cause clustering of transmembrane (TM) domains and stabilization of the transmembrane domains of the GPCR, termed the GPCR barrel due to their shape in the membrane (Venkatakrishnan et al, 2013). Ligand contact triggers movement of TM3 and causes conformational changes of the intracellular loops (ICL). In the Rhodopsin family, ICL2 contains an E/DRY motif near the boundary between ICL2 and TM3 (Rovati et al 2006). Mutations within this motif can result in constitutive GPCR signaling or impaired G-protein binding. Interestingly, other than the E/DRY motif, very little is known about how GPCRs associate with specific G-proteins. Chimeras resulting from swapping of ICL3 domains between GPCRs result in a switch in their G-protein selectivity (Kobilka et al 1988), suggesting that this domain imparts specificity for the downstream signaling pathway of the receptor.

The localization of a GPCR within a cell membrane can affect its ability to signal. Lateral movement of GPCRs within the plasma membrane is often restricted by the preferential

localization of GPCRs in specific lipid microenvironments (Allen et al 2007). Although GPCRs are usually shown as monomers, they can form oligomers (i.e. homo and heterodimers) within lipid rafts and are stabilized by interactions that are mediated through their TM domains (Ferre et al, 2014). Planar lipid rafts and caveolae both influence GPCR signaling by either excluding or recruiting G-proteins and their effectors. Because native GPCRs are not highly expressed in cells (McCusker 2007), lipid rafts concentrate GPCRs and their associated proteins to promote receptor dimerization and signaling. Caveolar-localized GPCRs are also subject to more rapid endocytosis, which represents another way to modulate GPCR signaling (Chini and Parenti 2004).

Signal propagation and the GPCR cycle

The presence of a ligand-receptor binding event must then be propagated and responded to by the cell itself. The effectors of GPCR activation are the heterotrimeric G-proteins G*α*, G*β*, and G*γ*. All organisms encode several types of each G-protein, and different combinations of these proteins into heterotrimers preferentially activates different signaling pathways. *Gα* is a GTPase, which catalyzes the hydrolysis of GTP to GDP. *Gα* is typically anchored in the membrane via N-terminal palmitolyation and G*α* can also be myristolyated (Vogler et al 2008). G*γ* is isoprenylated at its C-terminal CAAX motif (Higgins et al 1994). G*β* does not have any membrane-anchoring post-translational modifications. Instead, G*β* is tightly linked to G*γ* through hydrophobic interactions (Sondek et al 1996). The heterotrimeric complex can dock to an inactivated receptor or drift in the membrane, but once it encounters a ligand-bound GPCR, downstream signaling is initiated (**Fig. 1**). Activated GPCRs act as guanine nucleotide exchange factors (GEFs), and exchange GDP for GTP in the *Gα* subunit, which activates the protein. Upon GTP binding, G*α* changes its conformation, allowing it to separate from the Gβγ dimer. The subunits are then free to interact with downstream targets. When G*α* hydrolyzes GTP into GDP, it becomes inactivated, allowing G*α* to reassociate with Gβγ. This process represents a full GPCR G-protein cycle.

G*α* proteins are divided into four subclasses with each targeting a specific type of signaling cascade (Wettschurek et al 2005). G*αs* and G*αi/o* both regulate adenylate cyclase (AC). G*αs* stimulates AC activity, whereas G*αi/o* is inhibitory. The third subclass, G*αq/11* targets phospholipase C, which cleaves phosphatidylinositol 4,5-bisphosphate (PtdIns*P*2) into inositol triphosphate (IP3) and membrane-bound diacyl-glycerol (DAG). Finally, G*α12/13* activates Rho-GEFs, which in turn activate Rho. Previous models suggested that an individual GPCR interacts

with only one specific type of Gα, but it is now established that GPCRs are able to activate several Gα types, albeit with a marked preference for one (Cerione et al 1985).

Gαs is a weak GTPase, which slows the signaling cascade since new signaling information cannot be integrated (Kleuss et al 1994). To accelerate GTP hydrolysis, Gαs are targeted by the regulator of G-protein signaling (RGS) molecules (De Vries et al 2000). RGS proteins are somewhat promiscuous for Gαs, and several RGS proteins have been shown to bind to specific Gγs and prevent re-formation of the heterotrimeric complex (Witherow et al 2003). Conversely, activators of G-protein signaling (AGS) can act as GEFs for Gα to prolong signaling (Vogler et al 2008). Recent work has shown that Gα can be activated by non-receptor GEFs, which themselves are activated through associations with non-GPCR signaling pathways such as like the receptor tyrosine kinase (RTK) pathway (Garcia-Merces et al 2014). Some AGS proteins act as guanine-nucleotide-dissociation inhibitors (GDIs) (e.g. proteins containing a GoLoCo motif and GPR proteins) and cause Gα to remain in a GDP-bound state. These GDI-AGS proteins are termed G protein signaling modulators (GPSMs) and affect the amount of free G$\beta\gamma$ that can signal because they prevent re-association of the heterotrimeric protein complexes.

Compared to Gα, there are a fewer number of Gβ and Gγ genes in mammals (23 alpha, 5 beta, and 12 gamma genes). Therefore, there is a vast array of possible $\beta\gamma$ complexes, which each have different preferences for specific Gα subunits (Dingus et al 2005). Consequently, G$\beta\gamma$ can affect a wide range of ion channels and other signaling effectors, including pathways that are targeted by Gα, such as phospholipase C (Lau et al 2013). Because the G$\beta\gamma$ dimer was long considered to be a less important pathway component, the exact roles the different combinations of G$\beta\gamma$ dimers have on GPCR signaling remain to be elucidated.

Several mechanisms exist to attenuate GPCR signaling. The same conformational change in the GPCR that results in the release of the heterotrimeric complex results in residues becoming

accessible for phosphorylation by G-protein coupled receptor kinases (GRKs) (Palczewsk et al 1991). G$\beta\gamma$ is able to recruit a GRK to the GPCR, thus establishing a negative feedback loop (Luttrell et al 1999). GRKs usually phosphorylate GPCRs at serine or threonine residues in the ICL3 or the C-terminal tail of the activated receptor (Pitcher et al, 1998). This phosphorylation of the GPCR results in the binding of β-arrestin (Drake et al 2006), which then recruits clathrin and its adaptor AP-2, to internalize the GPCR. Several different fates await the GPCR following internalization and GPCRs are divided into two classes depending upon how strongly they maintain β-arrestin binding. Class A GPCRs lose β-arrestin following internalization (Oakley et al 2000) and can be dephosphorylated and recycled back to the cell-surface. An emerging concept in the field is the ability of GPCRs to maintain signaling once they are endocytosed (Mullershausen et al 2009). It is thought that by continuing to signal at endosomes instead of the plasma membrane, the physical distance between the GPCR and the nucleus is reduced, which results in more efficient signaling to transcriptional pathways (Tsvetnova et al 2014). Returning a GPCR back to the plasma membrane completes the GPCR cycle (**Fig. 1**). Class B GPCRs maintain β-arrestin binding. Here, GRK-mediated phosphorylation and binding to β-arrestin can stimulate the ubiquitination of GPCRs. Ubiquitinated receptors are then targeted to the lysozome for degradation. More recently, the so-called GPCR-associated sorting protein (GASP) family has been identified, which assists in the decision whether GPCRs are degraded or recycled (Bornert et al 2013; Simonin et al 2004). The amino acid sequence at the C-terminus of the GPCR is thought to direct the type of adaptor protein that binds the receptor, thus influencing the receptor's fate once internalized (Marchese et al, 2008).

GPCR families

The Rhodopsin-like family is the largest family of GPCRs in most organisms. Members of this family are well-known for the diversity of their ligands, which range from hormones to peptides to photons of light. Rhodopsin was first described as a light-sensitive compound in animals (Kuhne, 1877), and was cloned over a century later (Nathans and Hogness, 1983). This finding incited the explosion of subsequent GPCR research. Because the Rhodopsin-like family represents over 80% of human GPCRs (Fredrikkson and Schioth, 2005), it has been intensely studied for potential therapeutic benefits.

Another important GPCR familiar is the Secretin-like family; its main feature is its large N-terminal extra-cellular domain (ECD) (Watkins et al 2012), which is crucial for the recognition and binding of ligands, typically peptides or hormones. Historically, this family was named for the intestinal hormone secretin, which, in the early twentieth century, was the first hormone discovered (Bayliss and Starling, 1902). Its receptor was described nearly eighty years later (Jensen and Gardner, 1981; Chey and Chang, 2003).

Metabotropic glutamate-like (mGlu) receptors bind a diverse set of ligands, such as pheromones, amino acids, and calcium (Chun, 2012). Members of this family contain a large extracellular domain that forms a so-called venus flytrap (VFT) module (Bessis et al 2002). Upon ligand binding to one lobe of the VFT, the other lobe closes, introducing a conformational change that is transduced to the rest of the protein through a cysteine-rich region. mGlu receptors function as dimers, which are either covalently linked by disulfide bonds or by shared ion-binding. Compared to the other GPCR families, the mGlu family was discovered relatively late. Although glutamate was a known neurotransmitter, it was assumed to function solely through channels or ionotropic receptors, which themselves function as channels (Curtis and

Watkins, 1965). However, metabotropic receptors function as conventional GPCRs by binding a ligand and modulating that signal through the membrane.

The final family of GPCRs are the atypical GPCRs, which includes receptors such as Frizzled or adhesion GPCRs. Members of these families were initially thought to not signal primarily through heterotrimeric G-proteins (Tang et al, 2012). Members of the Frizzled family contain a cysteine-rich domain in their N-terminus that binds lipoglycoproteins of the Wingless (Wnt) family (Yang-Snyder, 1996). Upon ligand-binding, Frizzled family members primarily signal through the phosphoprotein Disheveled (Schulte and Bryja, 2007). GPCRs of the adhesion group often contain cadherin or integrin domains, and these receptors often contain auto-proteolytic activity (Krasnoperov et al, 1997). Their ligands include components of the extracellular matrix, such as collagen (Luo et al, 2011).

Ligand Identifiation

The most common way to characterize *Drosophila* GPCRs is through heterologous expression systems, such as Xenopus oocytes, HEK293 cells, or CHO cells. In this way, specific GPCRs can then be exposed to individual ligands to test for functionality. Heterologous expression systems are preferentially used to identify ligands for GPCRs (a process called "de-orphanizing") because it reduces the possibility of endogenous proteins triggering the receptor, as most GPCRs are quite specific for ligand recognition (Caers et al 2014). Many downstream effectors have not been singularly identified as coupling to a specific receptor, and instead output is typically measured through changes in calcium levels. Activation of *Gα*q releases calcium through the downstream effector PLC, and various calcium reporters can be used to track this change. However, changes in cAMP and RhoGEF activity are not as easily tracked. To circumvent this problem, GPCRs are usually expressed along with the promiscuous human *Gα*16, which readily couples with many GPCRs to cause changes in intracellular calcium. Whereas this system can identify ligand-receptor interactions, it does not identify the endogenous *Gα* that interacts with a GPCR. Moreover, as stated above, GPCRs are known to interact with more than one *Gα*, causing multiple output changes. Studies that have identified specific *Gα* coupling to a receptor have often been done using physical and genetic interaction data.

***Drosophila* GPCRs**

Drosophila encodes approximately two hundred GPCRs, which are categorized into the four types of families described above (Brody and Cravchik, 2000). Unsurprisingly, modulators of GPCR signaling, such as GRKs and arrestins, are also encoded in the *Drosophila* genome. One aspect of *Drosophila* GPCR signaling that makes it particularly appealing for study is the reduced number of G proteins – six alpha subunits, three beta, and two gamma (**Table 1) (**Katanayeva, 2010). Because flies have only eleven G-protein subunits, *Drosophila* offers a unique and advantageous system for studying GPCR signaling. However, compared to worms and mammals, the GPCR field in the fly is not as well-characterized. Nearly half of the *Drosophila* GPCRs are orphans, suggesting that the field is relatively wide open for new discoveries. Here, we present new insights into this field that may aid in future characterization of *Drosophila* GPCRs and we discuss the receptors' roles in embryonic development, which is also a rather unstudied side of GPCR signaling.

Phylogenetic analysis of *Drosophila* GPCRs

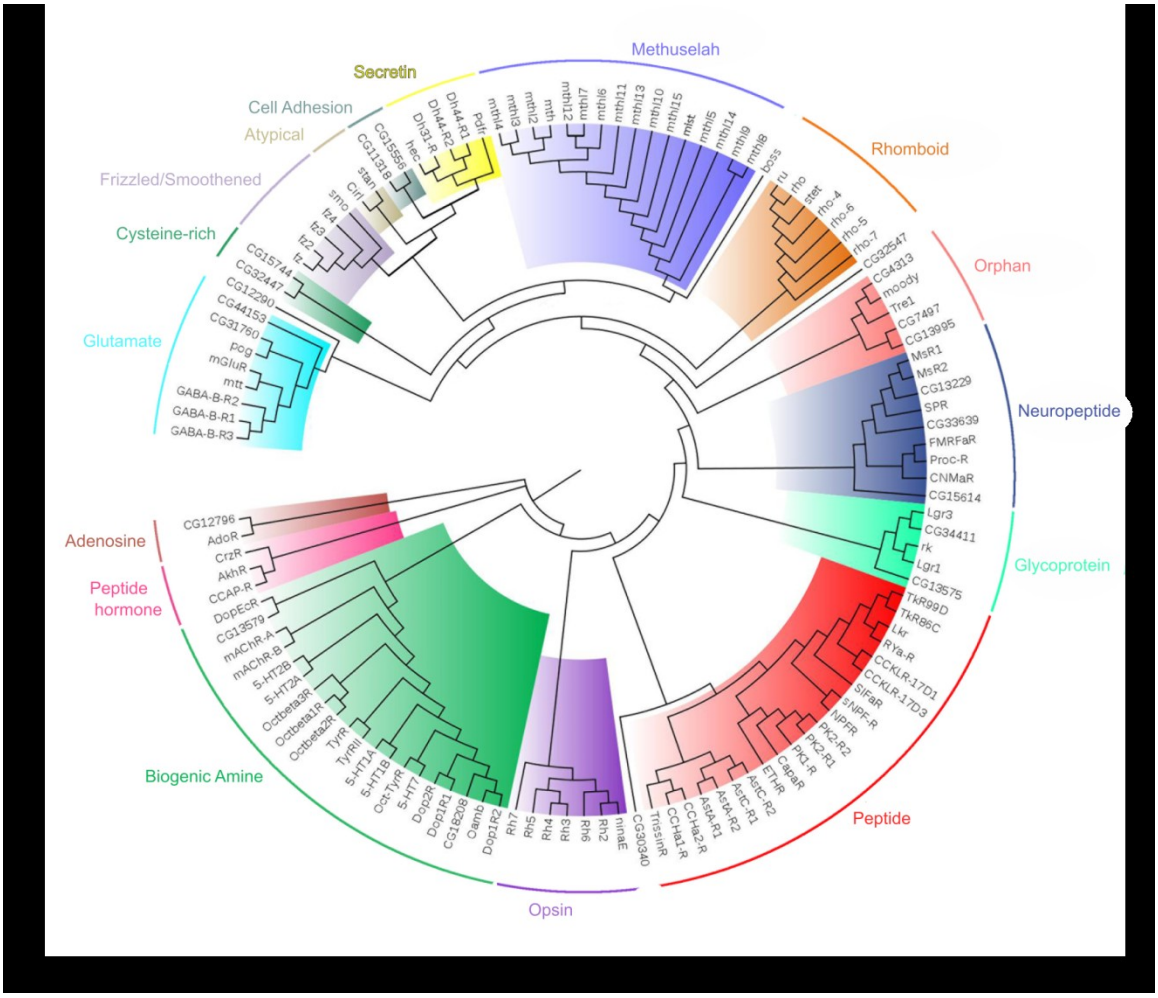
Using Clustal Omega with five combined iterations, we have created a phylogenetic tree using full-length sequences of all seven-transmembrane proteins that are encoded by the *Drosophila* genome (**Fig. 2, Supplemental Table 1**).For this analysis, we did not include odorant or gustatory receptors, which will be dicussed more below. We termed this GPCR set "NOG" GPCRS, since they are not odorant or gustatory. The NOG data set comprises 123 proteins, whose sequences were obtained from Flybase (www.flybase.org); for proteins with more than one isoform, only the first isoform listed was used. Historically, *Drosophila* GCPRs have been classified into four separate families: Rhodopsin-like, Secretin-like, metabotropic Glutamate-like, and Atypical. Based on our analysis, the NOG GPCRs separate into sixteen clades: Glutamate,

Table 1: Drosophila G-proteins.

Symbol	Name(s)	CG	Mammalian Output	Drosophila Expression Pattern
Gαi	G-iα65A, Gi, Gai	CG10060	inhibition of AC, ↓ cAMP	emb ubiquitious, emb brain, emb midgut, emb CNS, emb gonad, nurse cell, oocyte, maxillary palps
Gαo	G-oα47A, Go, Goα	CG2204	inhibition of AC, ↓ cAMP	maternally deposited, emb dorsal vessel primordium, visceral muscle primordium, emb brian, nurse cell, oocyte, maxillary palpus
Gαs	G-salpha60A, Gs, GSα	CG2835	activator of AC, ↑ cAMP	ubiquitous emb early, emb midgut primordium,emb hindgut primordium, emb visceral emb muscle primordium, adult brain cell body, maxillary palpus
Gαq	dgq, Galpha49b, Gqα	CG17759	activator of PLC, ↑ IP3/DAG, ↑ Ca2+	maternally deposited, emb head mesoderm, emb gut primordium, emb brain, emb Bolwig organ, adult eye, adult testis, maxillary palpus
Gαf	Gα73B, Galpha73B, Gfα	CG12232		emb midgut, oocyte
cta	concertina, conc	CG17678	activator of Rho via RhoGEFs	emb ubiquitous, oocyte, maxillary palpus
Gβ76C	Gβe, Gbeta76C	CG8770		emb Bolwig organ, adult eye, adult Johnston organ
Gβ13F	Gbeta13F	CG10545		maternally deposited, emb early ubiquitous,emb brain, emb midgut, emb hindgut, emb ventral nerve cord, adult ovary, adult brain cell, maxillary palpus
Gβ5	Gbeta5	CG10763		emb brain, emb midgut, emb ventral nerve cord, adult maxillary palpus
Gγ30A	Ggamma30A, Gγ, Gγe	CG3694		adult eye, maxillary palpus
Gγ1	Ggamma1, Gγ	CG8261		emb ubiquitous, emb brain, adult thorax, adult abdomen, ovary, oocyte, maxillary palpus

Figure 2: Rooted phylogenetic tree of *Drosophila* Not Odorant or Gustatory (NOG) GPCRs.

Full length sequences of GPCRs obtained from Flybase (www.flybase.org) were used to construct the tree. The names on correspond with the colored clades. Colors also correspond to those used in Supplemental Table 1.


Methuselah
Secretin
Cell Adhesion
Atypical
Frizzled/Smoothened
Cysteine-rich
Glutamate
Rhomboid
Orphan
Neuropeptide
Glycoprotein
Peptide
Opsin
Biogenic Amine
Peptide hormone
Adenosine
mth
mthl1
mthl2
mthl3
mthl4
mthl5
mthl6
mthl7
mthl8
mthl9
mthl10
mthl11
mthl12
mthl13
mthl14
mthl15
mtet
boss
ru
rho
stet
rho-4
rho-6
rho-5
rho-7
CG32547
CG43213
moody
Tre1
CG7497
CG13995
MsR1
MsR2
CG13229
SPR
CG33639
FMRFaR
Proc-R
CNMaR
CG15614
Lgr3
CG34411
rk
Lgr1
CG13575
TkR99D
TkR86C
Lkr
RYa-R
CCKLR-17D1
CCKLR-17D3
SIFaR
sNPF-R
NPFR
PK2-R2
PK2-R1
PK1-R
CapaR
ETHR
AstC-R2
AstA-R1
AstA-R2
CCHa2-R
CCHa1-R
Tnssin R
CG30340
ninaE
Rh2
Rh6
Rh3
Rh4
Rh5
Rh7
Dop1R2
Oamb
CG18208
Dop1R1
Dop2R
5-HT7
Oct-TyrR
5-HT1B
5-HT1A
TyrRII
TyrR
Octbeta2R
Octbeta1R
Octbeta3R
5-HT2A
5-HT2B
mAChR-B
mAChR-A
CG13579
DopEcR
CCAP-R
AkhR
CrzR
AdoR
CG12796
GABA-B-R3
GABA-B-R1
GABA-B-R2
mtt
mGluR
pog
CG31760
CG44153
CG12290
CG32447
CG15744
tz
tz2
tz3
tz4
smo
Ctrl
stan
CG11319
CG15556
hec
Dh31-R
Dh44-R2
Dh44-R1
Pdfr

Cysteine-rich, Frizzled/Smoothened, Atypical, Cell Adhesion, Secretin, Methuselah, Rhomboid, Orphan, Neuropeptide, Glycoprotein, Peptide, Opsin, Biogenic Amine, Peptide Hormone, and Adenosine. Four proteins (CG12290, Boss, CG32547, and CG30340) do not fit cleanly into any of these clades, but three of them are closely related to single clades: CG12290 and CG30340, which are uncharacterized genes, are closely related to the Glutamate receptors and Peptide receptors, respectively. Boss, an Atypical GPCR that acts as a ligand for a receptor tyrosine kinase, is closely related to the Methuselah receptors in our analysis. Both Boss and Methuselah receptors are known to have long extracellular domains, which may drive their close relationship in this phylogenetic tree.

Several interesting patterns emerge from the phylogenetic tree. Notably, the sixteen clades do not fit cleanly into the four previously described super families. Members of the metabotropic Glutamate-like family do not fall into one particular clade. In contrast, the Secretin-like and Atypical families are individual clades, which belong to a larger clade that, in addition, also contains methuselahs, cell adhesion GPCRs, Frizzled/Smoothened GPCRs, and cysteine-rich GPCRs. The classical Rhodpsin family is split into eight clades that cannot be rooted from a common point. As the Rhodpsin family is the largest GPCR family in all organisms, our tree represents a refinement of this large family into distinct sub-families.

This NOG-GPCR family tree might also point to potential roles for uncharacterized GPCRs or to their ligands. For instance, two GPCRs that initially were identified as orphan receptors (Brody and Cravchik, 2000) can now been sorted into families: CG33539 appears to be a member of the Neuropeptide family, whereas CG13579 belongs to the Biogenic Amine receptors clade and is closely related to DopEcR. Furthermore, CG12796 sorts into a small clade together with the Adenosine Receptor, suggesting these receptors might have similar roles. Finally, CG44153 is found to be a member of the Glutamate clade, whereas CG13575 sorts into the Glycoprotein

clade. Very little is known about these candidate genes (CGs), so this bioinformatic analysis suggests potential roles, ligands, and/or signaling pathways for each of these receptors.

We also constructed a phylogenetic tree that in addition to the NOG-GPCRs also includes the odorant and gustatory receptors (**Supplemental Figure 1**). This genome-wide comparison indicates that some of the GPCRs might not belong to clades they have been previously associated with. For example, Methuselah-like5 and Methuselah-like14 are not members of the Methuselah clade, but instead sort into a more distant clade, which includes the Secretin clade, Cell Adhesion clade, Atypical clade, and Rhomboid-7 (**Fig. 3A**).

The odorant receptors sort into one super-clade, with two notable exceptions. Or88a, one of only two known receptors that detect both male and female pheromones, belongs to the Frizzled clade (**Fig. 3B**). This unique localization within the phylogenetic tree is unexpected, and might reflect the special role this protein plays in the organism. Odorant co-receptor (Orco), the protein co-expressed with every odorant receptor protein, is also not a member of the odorant super-clade and instead is more closely related to Gr63a and Gr21a, the gustatory receptors for carbon dioxide (**Fi. 3C**). Interestingly, Orco is necessary for flies to sense carbon dioxide during flight, whereas Gr63a and Gr21a are not (Wasserman et at 2013). When odorant receptors are considered in a tree by themselves, Orco and Or88a are off-shoots of the same clade, but do not share a branch with any other protein (**Supplemental Figure 2, Supplemental Table 2**). Odorant receptors in insects have an inverse topology from typical GPCRs: the N-terminus is intracellular, whereas the C-terminus is extracellular (Benton et al 2006). Upon ligand binding, they activate Orco, which acts as an ion channel (Wicher 2008); this has led to the speculation that the *Drosophila* odorant receptors are not true GPCRs.

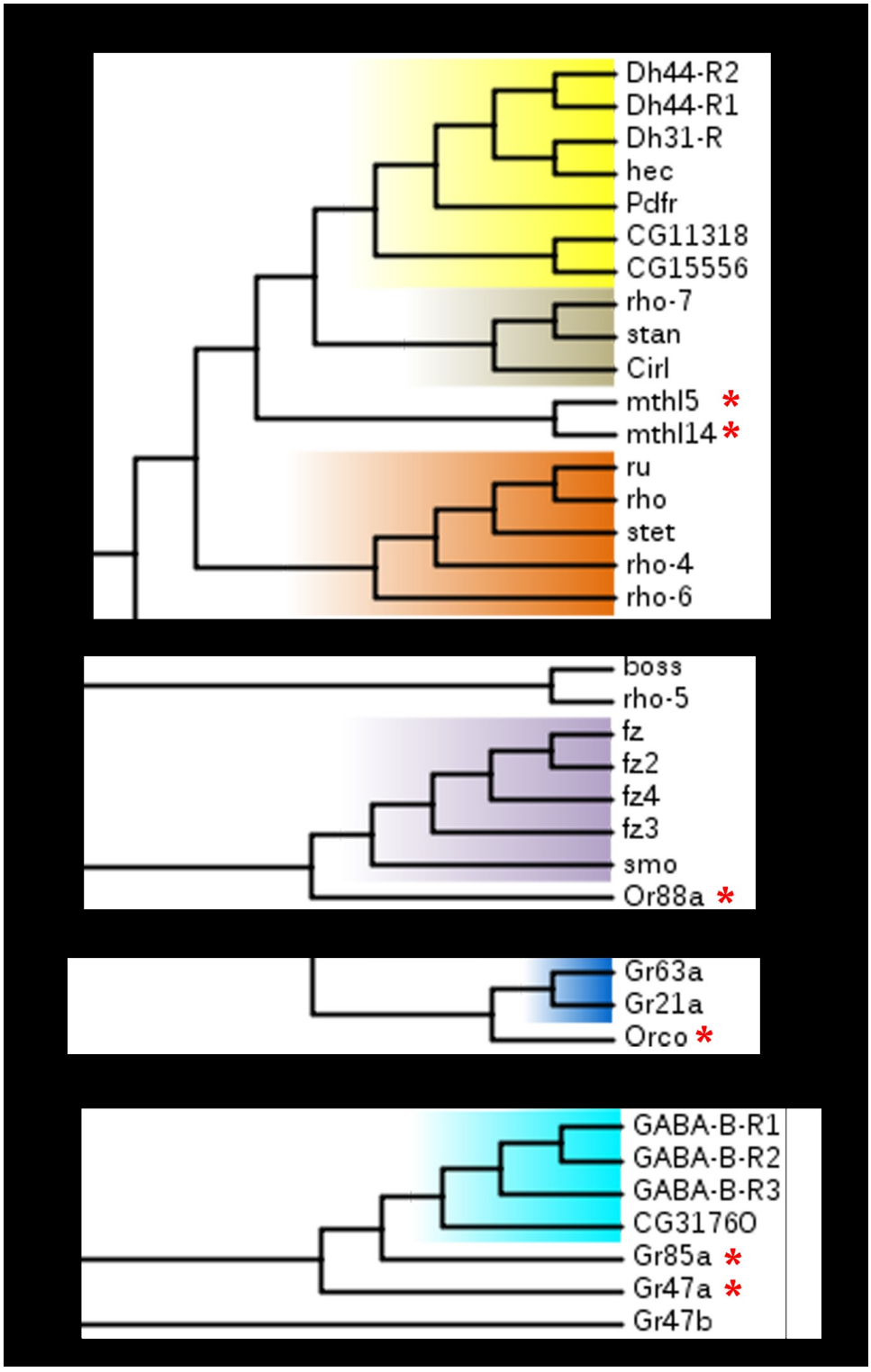
The *Drosophila* gustatory receptors separate into six clades when they are analyzed together with the other GPCRs (**Supplemental Figure 1**). Gustatory receptors are located on the

antenna, labellum, wing hairs and leg hairs, and sense a diverse array of compounds including sugars, bitter compounds, amino acids, and sex pheromones. The sweet-taste receptors (Gr64d, -64c, -64b, -61a, -64a, -5a, -64f, and -64e) all cluster together, as do the carbon dioxide receptors (Gr63a and Gr21a; see above). Forty of the sixty gustatory receptors cluster into a large super-clade. Of these, the two receptors for a female mating pheromone (Gr68a and Gr32a) cluster together, but the majority of the other receptors in this super-clade are largely uncharacterized. The remaining gustatory receptors belong to clades that are independent of the major gustatory clade. Of these, six gustatory receptors separate into a clade that is more closely related to that containing the odorant receptors and the NOG-GPCR superfamily. Among these is Gr59f, which clusters with the odorant receptors, suggesting that this protein may have a unique role. Three gustatory receptors cluster with the NOG GPCR superfamily, and interestingly, two of these, Gr85a and Gr47a, cluster with GABA receptors (**Fig. 3D**). Gustatory neurons that respond to sweet taste are also known to express GABA receptors, and loss of these receptors influences the perception of sugar (Chu et al 2014). Thus, these receptors might be involved in facilitating this interaction. In a phylogenetic tree that is only composed of the gustatory receptors, the receptors separate into the sweet clade, the carbon dioxide clade, and the female-mating pheromones clade as described above (**Supplemental Figure 3, Supplemental Table 3**).

GPCRS in *Drosophila* Development

Many of the characterized *Drosophila* GPCRs regulate adult behavior, but considerably less is known about their role in development. Before cellularization occurs, embryonic development in *Drosophila* is largely dictated by the action of diffusible gradients. Because the early embryo is essentially one cell with many nuclei, cell-to-cell communication is not necessary. However, as

Figure 3: GPCRs that localize in non-native clades in NOG, Odorant, and Gustatory phylogenetic tree. A. Mthl5 and Mthl14 do not cluster with the other members of the Methuselah family. B. Or88a clusters with the Frizzled family. C. Orco clusters with the carbon dioxide receptors Gr63a and Gr21. D. Gr85a and Gr47a cluster with the GABA family.

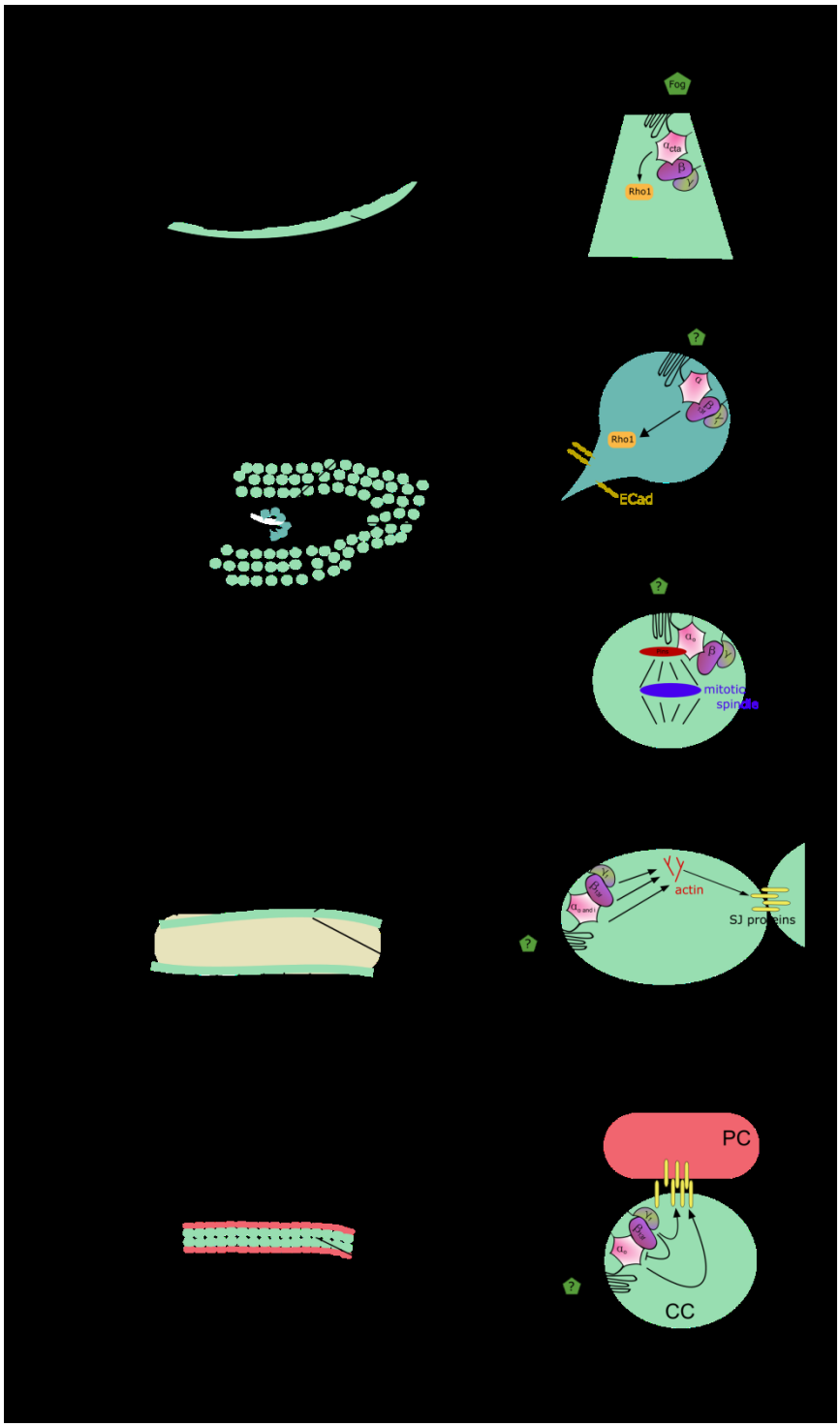

Dh44-R2
Dh44-R1
Dh31-R
hec
Pdfr
CG11318
CG15556
rho-7
stan
Cirl
mthl5 *
mthl14 *
ru
rho
stet
rho-4
rho-6
boss
rho-5
fz
fz2
fz4
fz3
smo
Or88a *
Gr63a
Gr21a
Orco *
GABA-B-R1
GABA-B-R2
GABA-B-R3
CG31760
Gr85a *
Gr47a *
Gr47b

development progresses, the activities of newly formed tissues and organs must be coordinated to from a viable organism. Although many GPCRs are expressed during embryonic stages, little is known with regard to their roles in specific developmental processes. Interestingly, only one G*β* (G*β13f*) and one G*γ* (G*γ1*) are expressed predominantly in embryos (**Table 1**).

One of the first major events of embryonic development is gastrulation. Concertina (a G*α*) and Fog (a secreted protein) were known to be necessary for this process, but the GPCR that linked these two molecules was unknown for over twenty years (Parks and Wieschaus, 1991; Costa et al 1994). In addition, the GEF Ric-8 was also identified as a crucial component for gastrulation (Peters and Rogers, 2013), further supporting the idea that a GPCR was involved in this process. Recently, Mist has been identified as a GPCR that coordinates these events and triggers the initial cell-shape changes (**Fig. 4A**; Manning et al 2013). The GPCR kinase Gprk2 acts downstream of Fog signaling and limits the cell-shape changes to ventral mesodermal cells (Fuse et al 2013), potentially by limiting Mist activity.

The atypical GPCRs of the Frizzled family are well-characterized for their role in segment polarity through the Wnt signaling pathway. Frizzled family members can also act as canonical GPCRs by signaling through G*o* (Katanev et al 2005), which in turn recruit Rab5, an important regulator of endocytosis, to the plasma membrane (Purvanov et al 2010). Endocytosis and morphogenesis are tightly linked, so this link between Go and Rab5 has potentially important implications for embryonic development. Frizzled and other proteins involved in planar polarity have also proven roles in tracheal development (Chung et al., 2009; Warrington et al 2013), but a specific role of Frizzled in GPCR signaling within the context of this developmental process has not been examined.

Figure 4: GPCRs in *Drosophila* development. A. The GPCR Mist binds to its ligand Fog in the early embryo; this activates the Gα Concertina (Cta) and its downstream effector Rho1 to mediate apical constriction of mesodermal cells during gastrulation. B. Tre1 has two roles in the developing embryo. In the germ cells, Tre1 activates Gβ13f and Gγ1 to relocalize Rho1 and E-cadherin (ECad) to the rear of the migrating cells. In neuroblast stem cells, Tre1 interacts with Gαo, which binds to Pins. Pins interacts with proteins that orient the mitotic spindle in such a way so that correct cell fates are established. C. Moody activates Gαi, Gαo, Gβ13f, and Gγ1 to alter the accumulation of actin in the surface glia, and affecting the organization of septate junction (SJ) proteins. D. An unknown GPCR is activated in cardial cells (CC), which activates Gαo, Gβ13f, and Gγ1. These proteins influence the localization of SJ proteins in CCs, which interact in trans with SJ proteins on pericardial cells (PC), thereby facilitating adhesion between these two types of cells to ensure proper cardiogenesis.


Fog
Rho1
Rho1
ECad
mitotic spindle
actin
SJ proteins
PC
CC

As in other organisms, embryonic germ cell migration is also directed by GPCR signaling in the fly. Loss of *Tre1* severely affects germ cell migration in the embryo (**Fig. 4B**; Kunwar et al 2003). *Tre1* was named for its phenotype: instead of migrating to the gonad, the germ cells remain trapped in the endoderm. The arginine in the DRY motif is necessary for Tre1 function (Kamps et al 2010). Tre1 was shown to influence distribution of Rho1 and E-cadherin in the germ cells, although the mechanism for this interaction remains unknown (Kunwar et al 2008). Tre1 also influences embryonic neuroblast stem cell divisions (Yoshiura et al 2012). Through interactions with its partners G*ao* and Pins (a GDI protein), Tre1 was shown to play a key role in the oriented cell divisions that establish neuroblast cell polarity and cell fate. Despite these known roles for Tre1 in controlling individual cell behavior, Tre1 remains an orphan receptor as its ligand is unknown.

There are also some insights into the role of *Drosophila* GPCRs in cell adhesion and boundary integrity. The Gα*o*-G*β13F*-G*γ1* complex mediates cardiomyocyte adhesion, and loss of G*γ1* causes mislocalization of septate junction proteins in the embryonic heart (**Fig. 4D**; Yi et al 2008). Septate junctions are also compromised in *moody* mutants, which functions in the late embryonic surface glia to form the blood brain barrier that insulates the nerve cord (**Fig. 4C**; Schwabe et al 2005). G*i*, G*o*, and Loco likely function in the same pathway, which is thought to directly affect the actin organization required for septate junction assembly. Like Tre1, Moody is also an orphan receptor, and these proteins belong to the same clade in our phylogenetic analysis. As many *Drosophila* GPCRs remain uncharacterized, exciting work remains to fully understand their potential roles in developmental processes.

Conclusions

In conclusion, GPCR signaling is vital to organismal life. GPCRs are able to translate an outside stimulus into a cellular response on a millisecond timescale, thus connecting a cell's behavior to its outer environment. GPCRs are involved in many key developmental events, but they are also involved in more refined events, such as discerning the difference between two closely-related scents. Ligands for GPCRs are immensely diverse, and many GPCRs remain orphans. Through phylogenetic analysis of the GPCRs encoded by the *Drosophila* genome, we hope to add new insights into the activities of this important family. Heterologous expression systems have historically been used to identify GPCR ligands. However, this method has limitations. Most notably, this method ignores spatiotemporal information, an aspect which is vital in living organisms. Most ligand-receptor pairs have been identified in adult organisms responding to behavior inputs such as odorants or mating pheromones. More nuanced analysis is needed to understand the role of GPCRs in developmental processes, where both components are genetically encoded and must be expressed at the right time and place to interact with one another. Bioinformatic analysis can point to potential shared or related ligands in clades where well-characterized and uncharacterized receptors cluster together. Our analysis suggests roles for several uncharacterized *Drosophila* GPCRs, which will hopefully spur further research about the exact ligand, pathway, and expression profile for each of these genes.

REFERENCES

Allen, J. A, Halverson-Tamboli, R. A, and Rasenick, M. M. (2007). Lipid raft microdomains and neurotransmitter signalling. *Nature Reviews: Neuroscience*, **8**, 128–40.

Attwood, T. K. and Findlay, J. B. C. (1994). Fingerprinting G-protein-coupled receptors. *Protein Engineering, Design and Selection*, **7**, 195–203.

Bayliss, W., and Starling, E.H. (1902). The mechanism of pancreatic secretion. *J Physiol*., **28**, 325-353.

Becker, E.L. Kanaho, Y., and Kermode, J.C. (1987). Nature and functioning of the pertussis toxin-sensitive G protein of neutorphils. *Biomed Pharmacother*, **41**, 289-297.

Benton, R., Sachse, S., Michnick, S. W., and Vosshall, L. B. (2006). Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo. *PLoS Biology*, **4**, e20.

Bessis, A-S., Rondard, P., Gaven, F., Brabet, I., Triballeau, N., Prezeau, L., Acher,F., and Pin, J-P. (2002). Closure of the Venus flytrap module of mGlu8 receptor and the activation process: Insights from mutations converting antagonists into agonists. *PNAS* **99**, 11097-11102.

Bornert, O., Moller, T.C., Boeuf, J., Candusso, M.P., Wagner, R., Martinez, K.L., Simonin, F., (2013). Identification of a novel protein-protein interaction motif mediating interaction of GPCR-associated sorting proteins with G protein-coupled receptors. *PLoS One*, **8**, e56336.

Brody, T., and Cravchik, A. (2000). Drosophila melanogaster G Protein – coupled Receptors. *J. Cell Biol.* **150**, 83–88.

Caers, J., Peymen, K., Suetens, N., Temmerman, L., Janssen, T., Schoofs, L., and Beets, I. (2014). Characterizatino of G protein-coupled receptors by a fluorescence-based calcium mobilization assay. *J. Vis. Exp.* **89**, e51516.

Cerione, R.A., Staniszewski, C., Benovic, J.L., Lefkowitz, R.J., Caron, M.G., Gierschik, P. Somers, R. Spiegel, A.M., Codina, J., and Birnbaumer, L. (1985). Specificity of the functional interactions of the beta-adrenergic receptor and rhodopsin with guanine nucleotide regulatory proteins reconstituted in phospholipid vesicles. *J. Biol. Chem.* **260**, 1493-1500.

Chang, T.M., and Chey, W.Y. (1980). Radioimmunoassay of secretin: A critical review and current status. *Dig. Dis. Sci*., **25**, 529-552.

Chini, B, and Parenti, M. (2004). G-protein coupled receptors in lipid rafts and caveolae: how, when, and why do they go there? *J Mol Endocrinol*., **32**, 325-38.

Chu, B., Chui, V., Mann, K., and Gordon, M. D. (2014). Presynaptic gain control drives sweet and bitter taste integration in Drosophila. *Current Biology*, **24**, 1978–84.

Chun, L., Zhang, W., and Liu, J. (2012). Structure and ligand recognition of class C GPCRs. *Acta Pharmacologica Sinica,* **33**, 312-323.

Chung, S.Y., Vining, M.S., Bradley, P.L., Chan, C.C., Wharton Jr, K.A., and Andrew, D.J. (2009). Serrano (Sano) functions with the planar cell polarity genes to control tracheal tube length. *PLoS Genetics*, **5**, e1000746.

Costa, M., Wilson, E. T., and Wieschaus, E. (1994). A Putative Cell Signal Encoded by the folded gastrulation Gene Coordinates Cell Shape Changes during Drosophila Gastrulation, *Cell*, **76**(6), 1075-1089.

Curtis, D.R. and Watkins, J.C. (1965). The pharmacology of amino acids related to gamma-aminobutyric acid. *Pharmacological Reviews*, **17**, 347-391.

De Vries, L., Zheng, B., Fischer, T., Elenko, E., and Farquhar, M.G. (2000). The regulatory of G protein signaling family. *Annu. Rev. Pharmacol. Toxicol.* **40**, 235-71.

Dingus, J., Wells, C. A., Campbell, L., Cleator, J. H., Robinson, K., and Hildebrandt, J. D. (2005). G Protein βγ Dimer Formation: Gβ and Gγ Differentially Determine Efficiency of in Vitro Dimer Formation. *Biochemistry*, **44**, 11882–11890.

Doitsidou, M., Reichman-Fried, M., Stebler, J., Koprunner, M., Dorries, J., Meyer, D. Esquerra, C.V., Leung, T., and Raz, E. (2002). Guidance of primordial germ cell migration by the chemokine SDF-1. *Cell*, **111**, 647-59.

Drake, M. T., Shenoy, S. K., and Lefkowitz, R. J. (2006). Trafficking of G protein-coupled receptors. *Circulation Research*, **99**, 570–82.

Ferré, S., Casadó, V., Devi, L. A., Filizola, M., Jockers, R., Lohse, M. J., Milligan G. Pin, J. Guitart, X. (2014). G Protein – Coupled Receptor Oligomerization Revisited : Functional and Pharmacological Perspectives. *Pharmacol. Rev.* **66**, 413–434.

Fishburn, C. S., Herzmark, P., Morales, J., Bourne, H. R., and Cell, H. R. M. B. (1999). Gβγ and Palmitate Target Newly Synthesized Gaz to the Plasma Membrane. *J. Biol. Chem.* **274**, 18793–18800.

Fredriksson, R., Lagerstrom, M.C., Lundin, L.G., and Schioth, H.B. (2003). The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272.

Fuse, N., Yu, F., and Hirose, S. (2013). Gprk2 adjusts Fog signaling to organize cell movements in Drosophila gastrulation. *Development*, **140**, 4246-55.

Fredriksson, R., and Schioth, H. B. (2005). The Repertoire of G-Protein – Coupled Receptors in Fully Sequenced Genomes. *Mol. Pharmacol.* **67**, 1414–1425.

Garcia-Marcos, M., Ghosh, P., and Farquhar, M. G. (2015). GIV/ Girdin transmits signals from multiple receptors by triggering trimeric G protein activation. *Journal of Biological Chemistry*, **290**, 6697-6704.

Higgins, J. B., and Caseys, P. J. (1994). In Vitro Processing of Recombinant G Protein γ Subunits, *J. Biol. Chem.* **269**, 9067–9073.

Houamed, K. M., Kuiper, J. L., Gilbert, T. L., Halderman, B.A., O'Hara, P. J., Mulvihill, E. R., Almers, W., and Hagen, F. S. (1991). Cloning, expression, and gene structure of a G protein-coupled glutamate receptor from rat brain. *Science,* **252**, 1318-1321.

Jensen, R.T., and Gardner, J.D. (1981). Identification and characterization of receptors for secretagogues on pancreatic acinar cells. *Fed Proc.* **40**, 2486-96.

Kamps, A. R., Pruitt, M. M., Herriges, J. C., and Coffman, C. R. (2010). An evolutionarily conserved arginine is essential for Tre1 G protein-coupled receptor function during germ cell migration in Drosophila melanogaster. *PloS One,* **5**, e11839.

Katanaev, V.L., Pnzielli, R., Semeriva, M., and Tomlinson, A. (2005). Trimeric G protein-dependent frizzled signaling in Drosophila. *Cell*, **120**, 111-22.

Katanayeva, N., Kopein, D., Portmann, R., Hess, D., and Katanaev, V. L. (2010). Competing activities of heterotrimeric G proteins in Drosophila wing maturation. *PloS One*, **5**, e12331.

Kleuss, C., Raw, A. S., Lee, E., Sprangt, S. R., and Gilman, A. G. (1994). Mechanism of GTP hydrolysis by G-protein a subunits, *PNAS,* **91**, 9828–9831.

Kobilka, B. (2012). The Structural Basis of G Protein Coupled Receptor Signaling. *Lecture, N.* 195–213.

Kobilka, B.K., Kobilka, T.S., Daniel, K., Regan, J.W., Caron, M.G., and Lefkowitz, R.J. (1988). Chimeric alpha 2-, beta 2-adrenergic receptors: delineation of domains involved in effector coupling and ligand binding specificity. *Science*, **240**, 1310-1316.

Kolakowski, L. F. (1994) GCRDB- A g-protein-coupled receptor database. *Journals of Receptors & Channels*, **2**, 1-7.

Krasnoperov, V. G., Bittner, M. A., Beavis, R., Kuang, Y., Salnikow, K. V, Chepurny, O. G., Little, A.R., Plotnikov, A.N., Wu, D., and Petrenko, A. G. (1997). α-Latrotoxin Stimulates Exocytosis by the Interaction with a Neuronal G-Protein-Coupled Receptor. *Neuron*, **18**, 925–937.

Kühne, W. (1877) Zur Photochemie der Netzhaut. Untersuch. Physiol. Instit. Univ. Heidelberg 1,1-14.

Kunwar, P.S., Starz-Gaiano, M., Bainton, R.J., Heberlein, U., and Lehmann, R. (2003). Tre1, a G protein-coupled receptor, directs transepithelial migration of Drosophila germ cells. *PLoS Biology*, **1**, E80.

Kunwar, P.S., Sano, H., Renault, A.D., Barbosa, V., Fuse, N., and Lehmann, R. (2008). Tre1 GPCR initiates germ cell transepithelial migration by regulating Drosophila melanogaster E-cadherin. *The Journal of Cell Biology*, 183, 157–68.

Lappano, R., and Maggiolini, M. (2011). G protein-coupled receptors: novel targets for drug discovery in cancer. *Nature Reviews. Drug Discovery*, **10**, 47–60.

Lau, W. W. I., Chan, A. S. L., Poon, L. S. W., Zhu, J., and Wong, Y. H. (2013). Gβγ -mediated activation of protein kinase D exhibits subunit specificity and requires Gβγ -responsive phospholipase C β isoforms. *Cell Communication and Signaling,* **11**, 1–17.

Lefkowitz, R. J. (2013). A brief history of G-protein coupled receptors (Nobel Lecture). *Angewandte Chemie (International Ed. in English)*, **52**, 6366–78.

Lemaire, K., Van de Velde, S., Van Dijck, P., and Thevelein, J. M. (2004). Glucose and sucrose act as agonist and mannose as antagonist ligands of the G protein-coupled receptor Gpr1 in the yeast Saccharomyces cerevisiae. *Molecular Cell*, **16**, 293–9.

Luo, R., Jeong, S.-J., Jin, Z., Strokes, N., Li, S., and Piao, X. (2011). G protein-coupled receptor 56 and collagen III, a receptor-ligand pair, regulates cortical development and lamination. *PNAS*, **108**, 12925–30.

Luttrell, L.M, Ferguson, S.S.G., Daaka, Y., Miller, W.E., Maudsley, S., Della Rocca, G.J., Lin, F.T., Kawakatsu, H., Owada, K., Luttrell, D.K. (1999). B-arrestin-dependent formation of the β2 adrenergic receptor-Src protein kinase complexes. *Science*, **283**, 655-661.

Manning, A.J., Peters, K.A, Peifer, M., and Rogers, S.L. (2013). Regulation of epithelial morphogenesis by the G protein-coupled receptor mist and its ligand fog. *Science Signaling*, **394**, ra98.

Marchese, A., and Paing, M.M., Temple, B.R.S., and Trejo, J. (2008). G protein-couple receptor sorting to endosome and lysosomes. *Annu Rev Pharmacol Toxicol*, **48**, 601-629.

McCusker, E. C., Bane, S. E., O'Malley, M. A., and Robinson, A. S. (2007). Heterologous GPCR expression: a bottleneck to obtaining crystal structures. *Biotechnology Progress*, **23**, 540–7.

Molyneaux, K. A. (2003). The chemokine SDF1/CXCL12 and its receptor CXCR4 regulate mouse germ cell migration and survival. *Development*, **130**, 4279–4286.

Mullershausen, F., Zecri, F., Cetin, C., Billich, A., Guerini, D., and Seuwen, K., (2009). Persistent signaling induced by FTY720-phosphate is mediated by internalized S1P1 receptors. *Nat. Chem. Biol.,* **5**, 428-434.

Nakayama, N., Miyajima, A., and Arai, K. (1985). Nucleotide sequences of STE2 and STE3 , cell type-specific sterile genes from Saccharomyces cerevisiae. *EMBO*, **4**, 2643–2648.

Nathans, J., and Hogness, D. S. (1983). Isolation, Sequence Analysis, and Intron-Exon Arrangement of the Gene Encoding Bovine Rhodopsin. *Cell*, **34,** 807–814.

Nordström, K. J. V, Sällman Almén, M., Edstam, M. M., Fredriksson, R., and Schiöth, H. B. (2011). Independent HHsearch, Needleman--Wunsch-based, and motif analyses reveal

the overall hierarchy for most of the G protein-coupled receptor families. *Molecular Biology and Evolution*, **28**, 2471–80.

Oakley, R. H., Laporte, S. A., Holt, J. A., Caron, M. G., and Barak, L. S. (2000). Differential affinities of visual arrestin, beta arrestin1, and beta arrestin2 for G protein-coupled receptors delineate two major classes of receptors. *J. Biol. Chem.* **275**, 17201–10.

Palczewskiss, K., Buczylkost, J., Kaplans, M. W., Polanss, A. S., and Crabb, J. W. (1991). Mechanism of Rhodopsin Kinase Activation. *J. Biol. Chem.* **266**, 12949–12955.

Parks, S., and Wieschaus, E. (1991). The Drosophila Gastrulation Encodes a Ga-like Protein Gene concertina encodes a Ga-like Protein. *Cell*, **64**(2), 447–458.

Peters, K.A. and Rogers, S.L. (2013). Drosophila Ric-8 interacts with the Ga12/13 subunit, Concertina, during activation of the Folded gastrulation pathway. *Mol. Biol. of the Cell*, **24**, 3460-71.

Pitcher, J. A., Freedman, N. J., and Lefkowitz, R. J. (1998). G protein coupled receptor kinases. *Annu. Rev. Biochem.* **67**, 653-92.

Prabhu, Y., and Eichinger, L. (2006). The Dictyostelium repertoire of seven transmembrane domain receptors. *European Journal of Cell Biology*, **85**, 937–46.

Purvanov, V., Koval, A., and Katanaev, V.L. (2010). A direct and functional interaction between go and rab5 during g protein coupled receptor signaling. *Sci. Signal.* **3**, ra65.

Rovati, G. E., and Neubig, R. R. (2007). The Highly Conserved DRY Motif of Class A G Protein-Coupled Receptors: Beyond the Ground State. *Mol. Pharmacol*. **71**, 959–964.

Schöneberg, T., Schulz, A., Biebermann, H., Hermsdorf, T., Römpler, H., and Sangkuhl, K. (2004). Mutant G-protein-coupled receptors as a cause of human diseases. *Pharmacology & Therapeutics*, **104**, 173–206.

Schulte, G., and Bryja, V. (2007). The Frizzled family of unconventional G-protein-coupled receptors. *Trends in Pharmacological Sciences*, **28**, 518–25.

Schwabe, T., Bainton, R.J., Fetter, R.D., Heberlein, U., and Gaul, U. (2005). GPCR signaling is required for blood-brain barrier formation in drosophila. *Cell*, **123**, 133–44.

Simonin, F., Karcher, P., Boeuf, J. J.-M., Matifas, A., and Kieffer, B. L. (2004). Identification of a novel family of G protein-coupled receptor associated sorting proteins. *Journal of Neurochemistry*, **89**, 766–75.

Sondek, J., Bohm, A., Lambright, D.G., Hamm, H. E., and Sigler, P. B. (1996). Crystal structure of a Gα protein βγ dimer at 2.1A resolution. *Nature*, **379,** 369-374.

Sugiyama, H., Ito, I., and Hirono, C. (1987). A new type of glutamate receptor linked to inositol phospholipid metabolism. *Nature*, **325**, 531-533.

Taddese, B., Upton, G. J. G., Bailey, G. R., Jordan, S. R. D., Abdulla, N. Y., Reeves, P. J., and Reynolds, C. A. (2014). Do plants contain g protein-coupled receptors? *Plant Physiology*, **164**, 287–307.

Takeda, S., Kadowaki, S., Haga, T., and Takaesu, H. (2002). Identification of G protein-coupled receptor genes from the human genome sequence. *FEBS Lett.* **520**, 97-101.

Tang, X., Wang, Y., Li, D., Luo, J., and Liu, M. (2012). Orphan G protein-coupled receptors (GPCRs): biological functions and potential drug targets. *Acta Pharmacologica Sinica*, **33**, 363–71.

Tsvetanova, N.G., and von Zastrow, M. (2014). Spatial encoding of cyclic AMP signaling specificity by GPCR endocytosis. *Nat. Chem. Biol.,* **10**, 1061-5.

Urano, D., Jones, J.C., Wang, H., Matthews, M., Bradford, W., Bennetzen, J.L., and Jones, A.M. (2012). G protein activation without a GEF in the plant kingdom. *PLoS Genetics*, **8**, e1002756.

Venkatakrishnan, A.J., Deupi, X., Lebon, G., Tate, C.G., Schertler, G.F., and Babu, M.M. (2013). Molecular signatures of G-protein-coupled receptors. *Nature*, **494**, 185–94.

Vögler, O., Barceló, J. M., Ribas, C., and Escribá, P. V. (2008). Membrane interactions of G proteins and other related proteins. *Biochimica et Biophysica Acta*, **1778**, 1640–52.

Vries, L. De, Zheng, B., Fischer, T., Elenko, E., and Farquhar, M.G. (2000). The regulator of G protein signaling family. *Annu. Rev. Pharmacol. Toxicol.* **40**, 235-71.

Warrington, S.J., Strutt, H., and Strutt, D. (2013). The Frizzled-dependent planar polarity pathway locally promotes E-Cadherin turnover via recruitment of RhoGEF2. *Development,* **140**, 1045-54.

Wasserman, S., Salomon, A., and Frye, M.A. (2013). Drosophila tracks carbon dioxide in flight. *Current Biology*, **23**, 301–6.

Watkins, H. A., Au, M., and Hay, D. L. (2012). The structure of secretin family GPCR peptide ligands: implications for receptor pharmacology and drug development. *Drug Discovery Today*, **17**, 1006–14.

Wettschureck, N., and Offermanns, S. (2005). Mammalian G proteins and their cell type specific functions. *Physiological Reviews*, **85**, 1159–204.

Wicher, D., Schafer R., Bauernfeind, R., Stensmyr, M. C., Heller, R., Heinemann, S. H., Hansson, B. S. (2008). Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. Nature, **452**, 1007–1011.

Witherow, D. S., and Slepak, V. Z. (2003). A novel kind of G protein heterodimer: the G beta5-RGS complex. *Receptors Channels*, **9***,* 205-212.

Yang-Snyder, J., Miller, J. R., Brown, J. D., Lai, C. J., and Moon, R. T. (1996). A frizzled homolog functions in a vertebrate Wnt signaling pathway. *Current Biology*, **6**, 1302–1306.

Yoshiura, S., Ohta, N., and Matsuzaki, F. (2012). Tre1 GPCR signaling orients stem cell divisions in the Drosophila central nervous system. *Developmental Cell*, **22**, 79–91.

Yi, P., Johnson, A.N., Han, Z., Wu, J., and Olsen, E.N. (2008). Heterotrimeric G proteins regulate a non-canonical function of septate junction proteins to maintain cardiac integrity in Drosophila. *Dev. Cell.* **15**, 704-13.

CHAPTER 3

The Role of the GPCR *Tre1* in Germ Cell and Salivary Gland Migration

ABSTRACT

Tre1 is a G-protein coupled receptor that is expressed in multiple tissues throughout embryogenesis, including the germ cells and the salivary gland. Previous work by several labs demonstrated a role for *Tre1* in germ cell migration (Kunwar et al 2003; Kamps et al 2010). These alleles, however, did not show defects in salivary gland migration. *In situ* analysis showed that the available mutant alleles were not complete nulls. Therefore, we made a clean *Tre1* null allele using homologous recombination. Germ cell migration was severely affected in the null allele, whereas the salivary gland still migrated correctly. This suggests that germ cells, as single cells, are sensitive to the loss of one signaling receptor, while the salivary gland, which migrates as an intact epithelium, uses compensatory factors to direct its migration. Closer examination of the salivary gland revealed that microtubule organization is disrupted in the *Tre1* knockout, a phenotype that is also observed during neuroblast division (Yoshiura et al, 2012). Over-expression of *Tre1* in the salivary gland causes defects in salivary gland migration, and mis-expression of *Tre1* in the trachea causes mis-migration of the tracheal dorsal branch. This indicates that migrating tissues are sensitive to levels of *Tre1*, which may enable them to react to *Tre1's* unknown ligand. We plan to further investigate the effect of over-expression of *Tre1* on the microtubule structure. Moreover, we want to eliminate other signaling pathways in combination with the *Tre1* knockout to tease apart the redundant pathways that direct salivary gland migration.

INTRODUCTION

Cell migration is an exquisitely intricate process common to many higher organisms. Variations in the signals driving cell movement, the distance cells travel, and whether cells migrate as individuals, as cell clusters or as intact epithelia can are all possible. Cell migration can be beneficial, as in development or wound healing, or detrimental, as in cancer metastasis. To begin to unravel the complexities inherent to collective cell migration, the *Drosophila* salivary gland provides a relatively simple model system for parsing the underlying molecular and cellular events.

The salivary gland begins as a placode of ~140 polarized columnar epithelial cells on the embryo surface that internalizes and moves dorsally. Dorsal movement persists until the gland reaches the gut mesoderm, where it reorients and begins posterior migration, which continues until the gland reaches its final correct position oriented along the anterior-posterior axis of the embryo. Since the salivary gland does not undergo cell division or cell death during morphogenesis, it provides an ideal system for studying the cell biological events underlying migration. Previous work has shown that the interaction of the salivary gland with several nearby tissues, including the visceral mesoderm, fat body, and somatic muscle, is vital for correct gland positioning (Vining et al, 2005). Integrin subunits expressed in both the salivary gland and in the mesoderm upon which the gland migrates are also necessary for proper gland placement (Bradley et al, 2003). Specific signaling pathways, such as Netrin (an attractant in the central nervous system and visceral mesoderm), Slit (a repellent from the midline glia), and Wnt signaling (repellents from the CNS and ventral nerve cord) are required for correct gland positioning (Harris and Beckendorf, 2007; Kolesnikov and Beckendorf, 2005). The migration defects associated with mutations for components of these pathways are often quite subtle and

never fully penetrant, suggesting that multiple additional pathways contribute to final correct salivary gland placement.

G protein coupled receptors (GPCRs) are the main conduit for cells to interact and respond to their outside environment. By spanning the membrane seven times, GPCRs are able to bind ligands extracellularly, triggering conformational changes intracellularly. These conformational changes activate the heterotrimeric G-proteins (Gα, Gβ, and Gγ), which then activate second messenger pathways that dictate the cell's response to the ligand. GPCRs play key roles in directed cell migration in systems as varied as *Dictyostelium* aggregation and mouse heart development (Prabhu et al, 2006; Waller-Evans et al 2010). In zebrafish, the GPCR CXCR4 is expressed in migrating primordial germ cells, and its ligand SDF1 is expressed in the same zone of migration (Doitsidou et al, 2002). Knockdown of either component results in disrupted germ cell migration. The homologues of these proteins were also found to play the same role in the mouse (Molyneaux, 2003).

In *Drosophila*, a GPCR was also found to direct germ cell migration. Using a small deletion combined with genomic rescue constructs containing a combination of WT and mutant versions of the two affected genes, Kunwar and others showed that *Tre1* (*Trapped in endoderm 1*) is essential for germ cell migration (Kunwar et al, 2003). Germ cells lacking *Tre1*remain stuck in the endoderm and do not migrate through the primordial midgut. The requirement for *Tre1* is cell autonomous, as germ cells mutant for *Tre1*do not reach the gonad in an otherwise wild-type background. This germ cell migration defect is linked to the failure to relocalize Gβ13f, Rho1, and E-Cadherin to the rear of the migrating germ cells (Kunwar 2008). This finding suggests that activation of Tre1 has the combined effect of re-localizing proteins to alter cell shape and polarities, and reducing adhesion, thus promoting migration. Notably, as with most GPCRs, Tre1 is an orphan receptor with no known ligand.

Tre1 was independently discovered to affect polarity in embryonic neuroblasts (Yoshiura et al 2012). In this system, Tre1 was shown to bind G*ao*, which when activated, binds the GoLoCo protein Pins. Pins, through its binding partner Inscuteable, interacts with Bazooka and other members of the Par complex. In turn, these proteins interact with components of the mitotic spindle to correctly orient cell division. Neuroblasts lacking *Tre1* deviate from the normal 90^{o} plane of division axis. Interestingly, the neuroblasts that show the *Tre1* defects recover later in development and are positioned correctly. Compensatory input from nearby glia provide additional unknown signals that ensure that neuroblasts are oriented correctly.

The BDGP in situ database revealed that *Tre1* is expressed in the salivary gland transiently, prior to and during the first stages of posterior gland migration. Our initial characterization of salivary gland phenotypes revealed only very subtle and low penetrant apical membrane irregularities but our studies also revealed that existing excision alleles of *Tre1* are not null. Therefore, a *Tre1* null allele, wherein nearly the entire open reading frame was replaced by the *white+* eye color marker, was created via homologous recombination to determine if the low penetrance of salivary gland migration defects was due to residual *Tre1* function. Here, we characterize the defects associated with complete loss of *Tre1 Tre1*I in both germ cell and salivary gland migration, as well as in other tissues that express *Tre1*. We also created both an untagged and C-terminal GFP tagged version of Tre1 that allows for overexpression of *Tre1* and for protein localization.

MATERIALS AND METHODS

Fly Strains

The *Tre1* knockout line was made by Deborah Andrew using homologous recombination, as described by Gong and Golic (2003). *ΔEP5* and *ΔEP19* were obtained from Kunio Isono (Tohoku University, Sendai, Japan). UAS-tre1 and UAS-tre1-GFP were generated by Gateway cloning (Invitrogen, Carlsbad, CA) into the pTW and pTWG UAS vectors, respectively. These constructs were then injected into *w1118* flies by Rainbow Transgenic Flies, Inc (Camarillo, CA). *nos*-Gal4 was provided by Ruth Lehmann (New York University School of Medicine). MS1096-Gal4 was obtained through the Bloomington Stock Center (Bloomington, Indiana). jal^{5929} was made by Deborah Andrew (Andrew et al 1994). *btl*-Gal4 was created by Shiga et al (1996). *fkh*-Gal4 was created by our laboratory (Henderson and Andrew, 2000). Homozygous lethal lines were balanced over FM7-Ftz-LacZ, Cyo-Ftz-LacZ, or TM6-Ubx-LacZ embryos, and mutant chromosomes were identified by lacZ staining.

Tre1 antibody generation

DNA fragment of either the first 102 basepairs (corresponding to N-terminal extracellular 36 amino acids) or the last 201 basepairs (corresponding C-terminal intracellular 67 amino acids). These fragments were cloned in-frame into the pGex-6p-1 vector using XhoI and EcoRI sites (GE Healthcare Bio-Sciences, Pittsburgh, PA) and expressed in *E Coli*. Protein was induced during log growth phase with 0.1M IPTG and grown over-night at 16^{o}C. Inclusion body preps were performed to isolate the induced protein and injected into either rats or rabbits (Covance).

Immunohistochemistry and in situ hybridizations

In situ hybridization and immunohistochemistry were performed as previously described (Lehmann and Tautz, 1994; Reuter et al., 1990). The *Tre1 in situ* probe was generated from the RE0771 cDNA (Drosophila Genomics Resource Center). The *cg25c* cDNA was obtained from

DGRC. Antibody concentrations used in this study are as follows: RbαVasa (1:200; Santa Cruz Biotechnology), RbαGooseberry (undiluted; R. Holmgren), RbαForkhead (1:2000; a gift from S. Beckendorf, Berkeley, CA, USA), RtαCrebA (1:1,000) (Andrew et al., 1997), RbαSAS (1:500; a gift from D. Cavener, Penn State University, PA, USA), mααSpectrin (1:2; Developmental Studies Hybridoma Bank, Iowa City, IA), mαCrumbs (1:100; DSHB), mαFasIII (1:100; DSHB), mαTubulin (1:10, DSHB), RbαGFP (1:500; Invitrogen) and mαβgalactosidase (Promega, 1:10,000). All secondary antibodies (Vector Labs and Molecular Probes) were used at a 1:500 dilution. HRP images were developed using the Vectastain ABC kit (Vector Laboratories). Confocal images were obtained using a Zeiss LSM 510 or 700 Meta confocal microscope, using a Plan-Neofluor 63x and 100x, 1.3 oil objective and the Zeiss LSM software. All other images were obtained using a Zeiss Axiophot microscope configured with a Nikon Coolpix 4500 digital camera or a Janoptik ProgResC14 Plus opitical imaging system. Images were taken using a Plan-Neofluor 20x, 0.50 objective. Images were rotated and cropped using Inkscape, an open-source vector graphic editor. All images were obtained at room temperature.

RESULTS

Existing alleles of Tre1 are not completely nulls

Tre1 is an X-chromosome encoded G-protein coupled receptor (GPCR) expressed dynamically throughout the *Drosophila* embryo. *Tre1* is expressed in the head mesoderm, trunk mesoderm, and endoderm early (**Figure 1A**), and is also expressed in the salivary gland and fat body starting at stage 11. From stage 14 onward, *Tre1* is expressed in the proventriculus, in a small subset of midgut cells, in the central nervous system, the hindgut, and in pericardial cells. Expression of *Tre1* in the salivary gland occurs at a particularly interesting time with regards to organ migration. Expression of *Tre1* is first observed during internalization (st 11) and continues through the early stages of posterior migration; expression no longer detected by stage 14. Because GPCRs are known to play a vital role in directional sensing and migration, we set out to determine the role of *Tre1* in salivary gland migration and to study its function in the other various tissues.

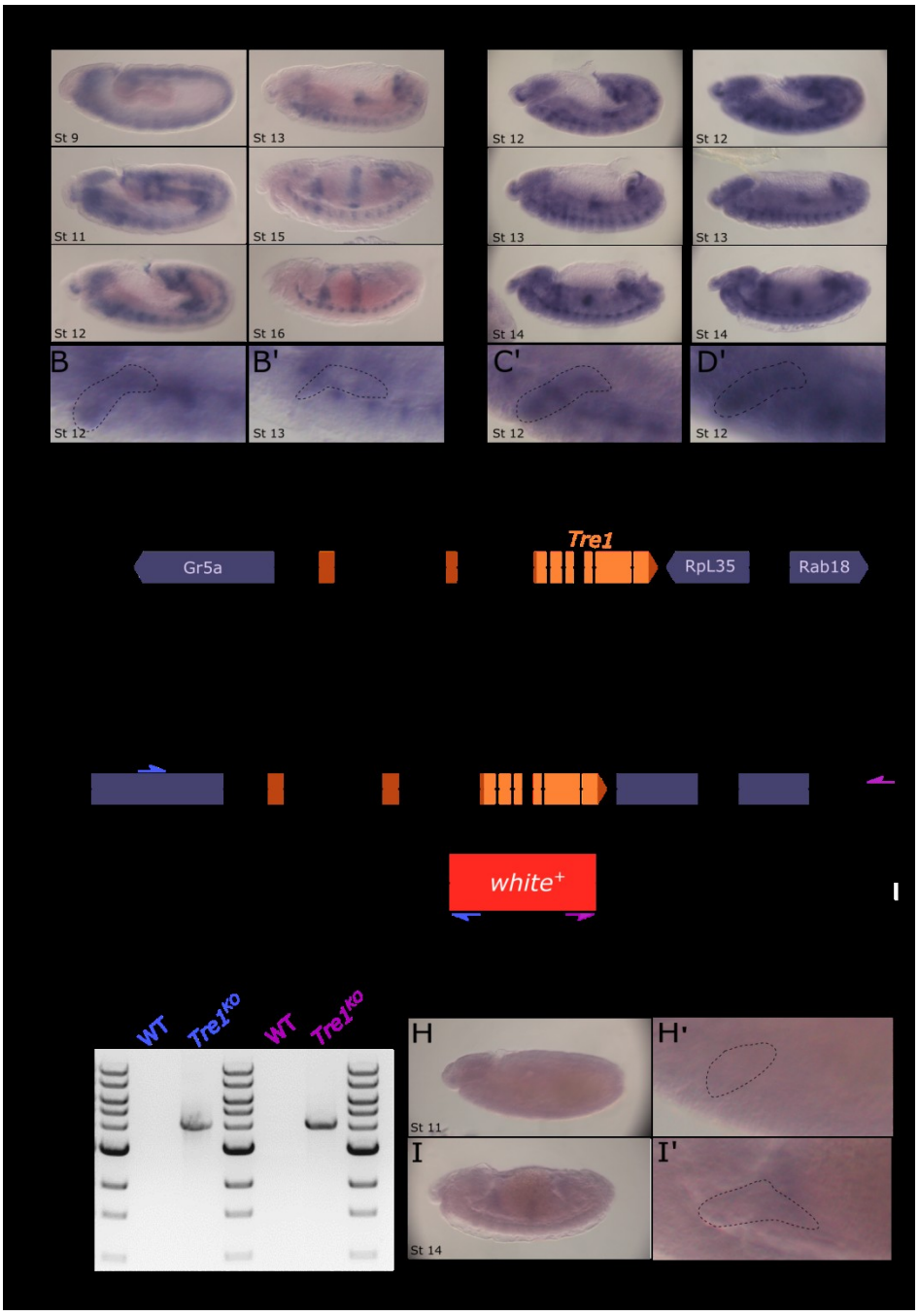
The two existing P-element excision alleles of *Tre1* were obtained, *ΔEP19* and *ΔEP5* (Ueno et al., 2011; **Figure 1B**). *ΔEP19* is a small deletion that removes the 5'UTR of *Tre1* including part of the first exon. *ΔEP5* is a larger deletion that also removes the 5'UTR and the entire first exon. Notably, neither of these alleles remove the start codon of *Tre1*, and an alternative transcription start site closer to the *Tre1* ORF has subsequently been discovered based on cDNA library screening (Flybase). Unfortunately, there are no deficiencies that remove *Tre1*, potentially because the essential ribosomal protein gene *rpl35* is located less than one kilobase from *Tre1* and heterozygosity for this gene may be problematic .

Previous work has shown that *Tre1* is required when the germ cells migrate through the endoderm toward the somatic gonad (Kunwar et al. 2003). This phenotype was observed using the *ΔEP5* allele mentioned above. When these lines were evaluated for their affects on salivary

Figure 1: A *Tre1* null allele was created because existing alleles still expressed *Tre1* transcript.

A-B''. *Tre1* transcript pattern in wild-type embryos. An in situ probe was made from full-length RE0771 cDNA, which is from the *Tre1-A* isoform. **A**. At stage 9, *Tre1* is expressed in the endoderm, mesoderm, and head. At stage 11 and 12, *Tre1* expression continues in those tissues and is found in the salivary gland and fat body. At stage 13, expression in the salivary gland is lost and expression in the midgut and hindgut appears. At stage 15, staining in the proventriculus, midgut, hindgut and CNS is visible. At stage 16, staining in the cardiac mesoderm is visible. **B', B''**. Inset showing salivary gland (outlined) *Tre1* expression is observed in stage 12 and is gone by stage 13. **C-D'.** *ΔEP19* and *ΔEP5* express *Tre1*. B' and C' are insets of stage 12 embryos showing *Tre1* expression. **E.** The *Tre1* genomic region on the X-chromosome. The 5' end of *tre1* is closer to the telomere and 3' end is closer to the centromere, so the orientation of the gene in this diagram is the opposite of what is shown on Flybase. The untranslated 5' and 3' UTRs of *Tre1* tanscript are shown in brown, whereas the open reading frame is shown in orange. There are three isoforms of *Tre1*: Isoforms A and C, which have the same transcription start site, and isoform B, which has a transcription start site closer to the coding exons. *ΔEP19* and *ΔEP5* were created by excising the p-element EP496; Both disrupt the gene encoding the gustatory receptor for trehalose, Gr5A. *ΔEP19* does not disrupt the first exon of *Tre1*, whereas *ΔEP5* deletes the first exon and half of the first intron of the A and C isoforms, but does not affect *Tre1-B*. **F.** Creation of a *Tre1* null allele via homologous recombination. Each homology arm is approximately four kilobases (kb) long. The *white+* replaces a region spanning from approximately 540nt upstream of the ATG up to the residues encoding the last twenty three residues of Tre1. Half-arrows indicate the primers used for PCR analysis in F. **G.** PCR analysis of the *Tre1* null allele. Primers were designed to sit either just outside the region included in the clone designed for recombination or within the *white+* gene. The resulting product is expected

to be just over four kb. **H-I'**. *Tre1 in situ* in the *Tre1*^{KO} allele at stage 11 (H) and stage 14 (I). H' and I' show the outlined salivary gland at both stages.


St 9
St 13
St 11
St 15
St 12
St 16
B
St 12
B'
St 13
St 12
St 12
St 13
St 13
St 14
St 14
C'
St 12
D'
St 12
Tre1
Gr5a
RpL35
Rab18
white+
WT
Tre1KO
WT
Tre1KO
H
St 11
H'
I
St 14
I'

gland migration, only modest irregularities were observed. The low penetrance of salivary gland irregularities was somewhat surprising given the severe effects loss of *Tre1* had on germ cell migration. These differences could be attributed to several potential factors: (1) single cells may be more sensitive to changes in signaling than groups of cells, (2) it may simply be more obvious when a single cell mismigrates than when a single cell in an entire epithelium mismigrates, (3) the salivary gland may have and use more compensatory cues to direct migration, or (4) the existing alleles may not impact salivary gland expression or function of *Tre1*. Indeed, neither excision allele appears to be null for *Tre1* function in our hands (**Figure 1C,D**). Both *ΔEP19* and *ΔEP5* express variable levels of *Tre1* transcript, including in the salivary gland, thus complicating our analysis of its function in the salivary gland and other tissues.

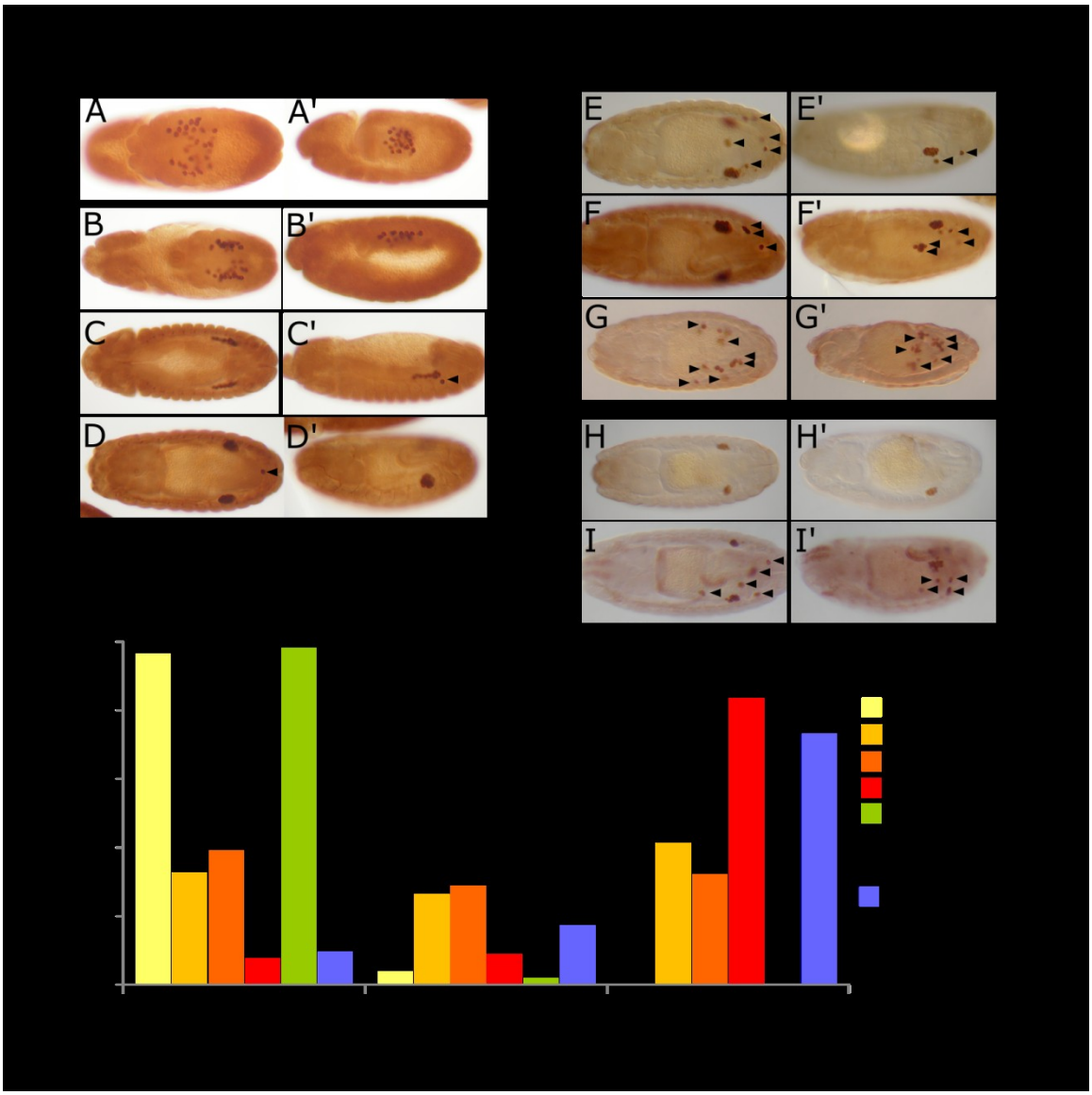
Creation of a Tre1 null allele

To create a *Tre1* null allele, we used homologous recombination to replace the *Tre1* ORF with that of the *white+* eye color gene (**Figure 1E**). Over nine hundred potential *Tre1* knockout lines were screened for replacement of the *Tre1* ORF with *white+*. Although several lines were obtained in which recombination appeared to have occurred on one side or the other of the coding region, only a single line was a clean replacement; this allele will be referred to as $Tre1^{KO}$. The complete knockout was verified by PCR and by *in situ* hybridization to assay for *Tre1* transcripts (**Figure 1F-H'**).

Germ cell migration is severely impaired in the Tre^{KO} flies

We first examined germ cell migration in the $Tre1^{KO}$. In wild-type embryos, germ cells are clustered in the posterior midgut during embryonic stage 9. During stage 10, they migrate through the endoderm as single cells (**Figure 2A,A'**). The germ cells then associate with

Figure 2: Germ cell migration is severely disrupted in the *Tre1*^{KO} allele. A-D'. Germ cell migration in wild-type embryos. Germ cells are stained with Vasa. Dorsal views are shown in A-D and lateral views are shown in A'-D'. At stage 11, germ cells have started to move out of the midgut endoderm (A') and begin to separate away from the midline into two bilateral groups of germ cells (A). At stage 12, the germ cells move anteriorly with germ band retraction. At stage 13, the germ cells are clustered in a line along the anterior-posterior axis. By stage 15, the germ cells are clustered tightly into the gonad. Arrowheads point out germ cells that have not coalesced with the other germ cells. **E-G'.** Germ cell migration at stage 15 in *Tre1* mutant alleles. E-G are dorsal views, E'-G' are lateral views. **H-I'.** Germ cell migration in *nanos* driven *Tre1* over-expression. **J.** Histogram of germ cell migration in *Tre1* alleles. Numbers of mismigrating germ cells were binned into zero to five, six to ten, or over eleven germ cells.

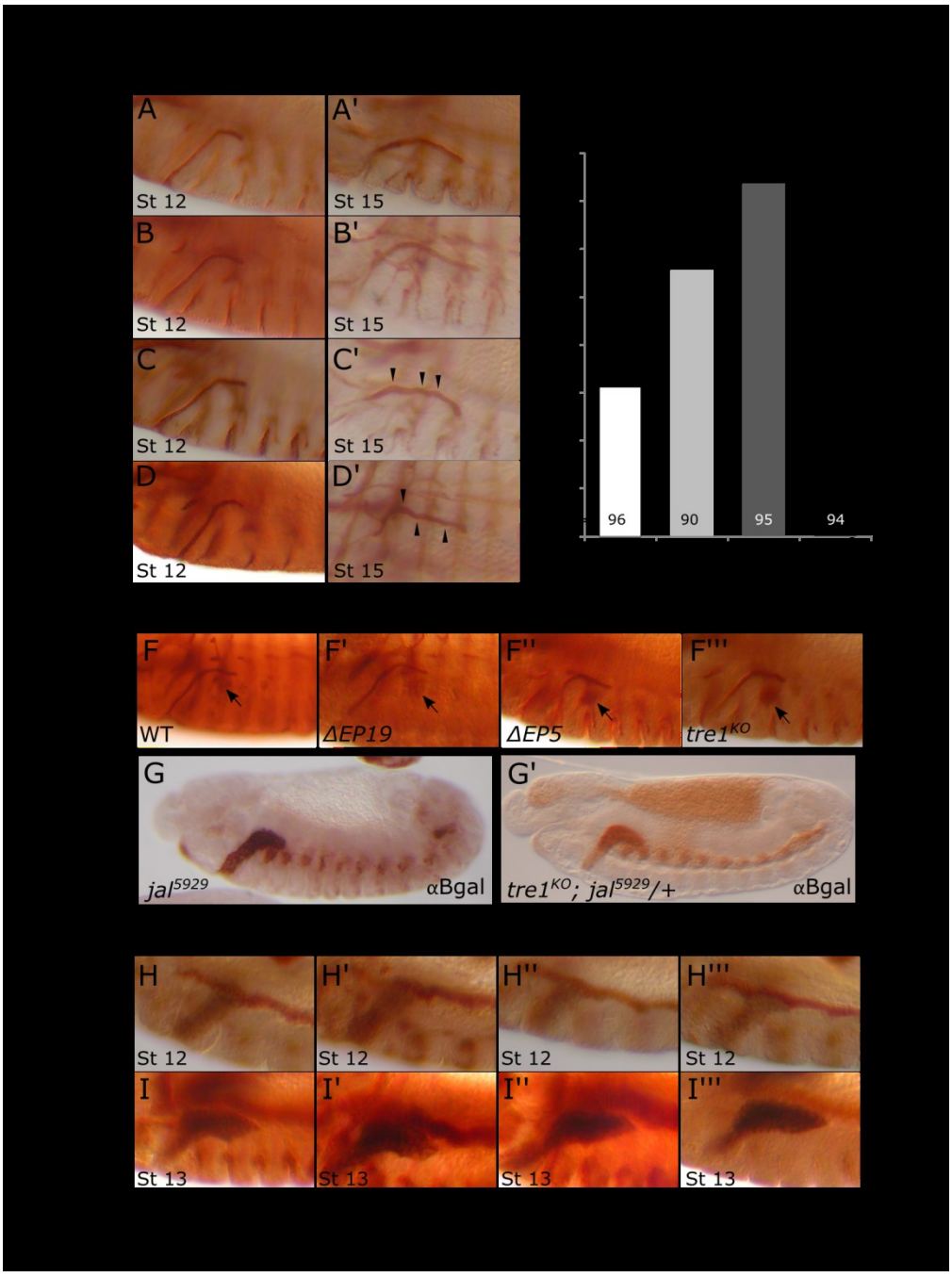

A
A'
B
B'
C
C'
D
D'
E
E'
F
F'
G
G'
H
H'
I
I'

mesodermal cells and migrate anteriorly through stage 13 (**Figure 2C**). By stage 14, the germ cells have coalesced into the gonad, where they remain for the remainder of embryogenesis (Figure 2D). Both *ΔEP19* and *ΔEP5* have germ cell migration defects (**Figure 2E-F'**), where a number of germ cells fail to reach the gonad. The $Tre1^{KO}$ germ cell migration phenotype is, however, far more severe than the previously described excision alleles. The $Tre1^{KO}$ germ cells migrate out of the midgut, but seem unable to navigate to the gonad. Germ cells in the *Tre1* null are found throughout the posterior end of the embryo (**Figure 2G-G'**). Mismigrating germ cells were counted in WT and *Tre1* mutant animals and categorized into groups depending on the total number of mismigrating cells: zero to five, six to ten, or over eleven germ cells mismigrating (**Figure 2J**). Over ninety percent of wild-type embryos had fewer than five mismigrating germ cells, whereas the majority of $Tre1^{KO}$ embryos had over eleven mismigrating germ cells. *ΔEP19* and *ΔEP5* had an intermediate phenotype. From these data, we conclude that complete loss of *Tre1* results in more severe and more penetrant germ cell migration defects that either of the previously characterized excision alleles.

Salivary gland migration is not affected with loss of Tre1

Salivary gland migration was only mildly affected with *Tre1* loss (**Figure 3A-D'**). Crb staining was used to evaluate the apical surface membrane of the migrating salivary gland. In wild-type embryos, the apical membrane is even and smooth at the beginning of migration (stage 12) and at the end of gland placement (stage 15). Only minor irregularities in the apical membrane are seen in *ΔEP5* and $Tre1^{KO}$, and these irregularities occurred with relatively low penetrance (**Figure 3E**). Thus, the failure to observe defects in the salivary gland versus the germ cells is likely due to the relative insensitivity of the polarized collective of salivary glands to signal perturbation versus single migrating germ cells or to other compensatory pathways guiding migration.

Figure 3: Salivary gland migration is not affected in *Tre1* mutant alleles. A-D'. Crb staining showing the apical membrane of the salivary gland at stage 12 (A-D) and stage 15 (A'-D'). Arrowheads point out slight apical membrane irregularities. **E.** Quantification of apical membrane irregularities in the *Tre1* mutant alleles. **F-F'''.** Gooseberry (arrow) and Crb staining showing the relationship between stage 12 salivary glands and the fat body. **G-G'.** jal^{5929} reporter expression in wild-type (F) or $Tre1^{KO}$. **H-I'''.** SG2 (ER) and FasIII (cVM) staining at stage 12 (H-H''') and stage 13 (I-I'''). At stage 12, the SG and cVM are very close to one another. At stage 13, the distal tip of the gland begins to separate from the cVM.


A
St 12
A'
St 15
B
St 12
B'
St 15
C
St 12
C'
St 15
D
St 12
D'
St 15
96
90
95
94
F
WT
F'
ΔEP19
F''
ΔEP5
F'''
tre1KO
G
jal5929
αBgal
G'
tre1KO; jal5929/+
αBgal
H
St 12
H'
St 12
H''
St 12
H'''
St 12
I
St 13
I'
St 13
I''
St 13
I'''
St 13

Salivary glands migrate in close contact with a number of different tissues, including the fat body and the visceral mesoderm. Thus, we examined the positional relationship between the fat body and the salivary gland during early stages of posterior migration. Co-staining with Crb and Gooseberry (Gsb) reveals the salivary gland lumen and the clusters of thoracic fat body cells (**Figure 3F-F'''**). During stage 13, the salivary gland is completing its posterior turn, just dorsal to these fat body clusters. In both wild-type embryos and in the *Tre1* mutant alleles, salivary gland position relative to the fat body appears entirely normal. The relationship between the fat body and the salivary gland was also examined using the jal^{5929} beta-gal insertion line, which marks nuclei in both the salivary gland and the fat body clusters (**Figure 3G,G'**). Using these markers, no difference was observed in the positioning of the salivary gland with respect to the fat body in wild-type versus $Tre1^{KO}$ mutants.

The relationship between the salivary gland and the gut mesoderm was also examined. During posterior salivary gland migration, the dorsal portion of the gland is in direct contact with the circular visceral mesoderm (cVM; Vining et al 2005). This contact is disrupted at the beginning of stage 14/end of stage 13when the longitudinal visceral mesoderm (lVM), which comes from a very posterior position in the embryo, migrates between the cVM and the gland. We examined the relative position of the salivary gland with respect to the cVM by staining with the salivary gland specific endoplasmic reticulum marker SG2 and with FasIII, which marks the cVM (**Figure 3H-I'''**). In wild-type stage 12 embryos, the distal portion of the dorsal side of the gland is tightly adhered to the cVM. By stage 13, the distal tip of the salivary gland begins to separate from the cVM as the lVM cells migrate in between them. The same physical relationship between the salivary gland and cVM is seen in the *Tre1* null mutant embryos. Thus, even when examined in combination with markers that highlight tissues the salivary gland

contacts during early stages of its migration, we see no overt changes in salivary gland migration with either partial or complete loss of *Tre1*.

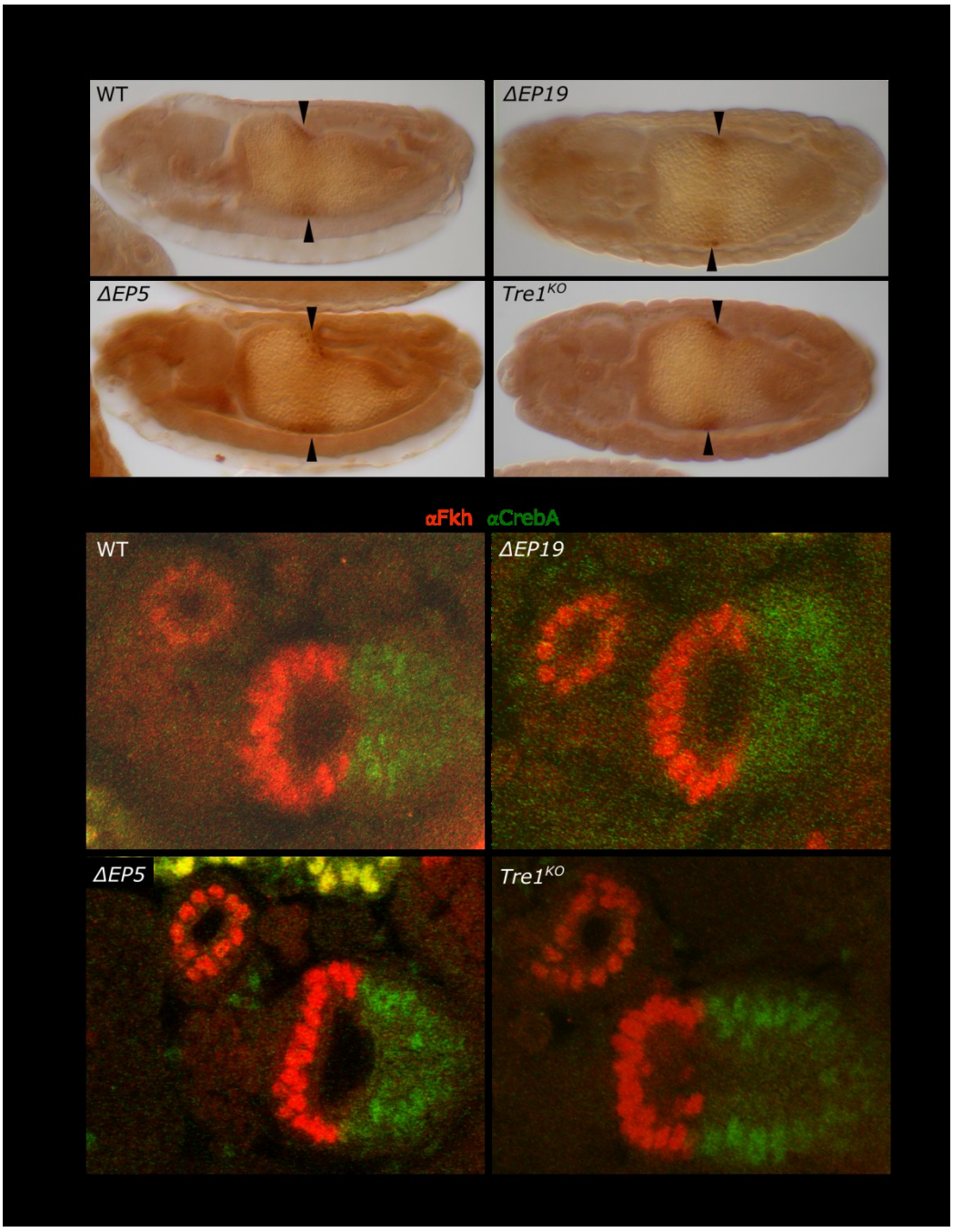
Other tissues are not affected in Tre1^{KO} flies

Tre1 is also expressed in the midgut and the proventriculus. We assayed for potential defects in the midgut by staining for Labial, which marks the cuprophilic cells that form near the second midgut constriction, which is the first constriction to form (Tremml and Bienz, 1992). At stage 15, Labial staining is present at the gut constriction and in a light band around the gut (**Figure 4A**). Similar staining is seen in *ΔEP19, ΔEP5*, and *Tre1^{KO}*. Therefore, we conclude that the midgut forms normally in the absence of *Tre1*. The proventriculus was examined by staining for Fork head, a transcription factor that is expressed in the proventricular ectodermal cell nuclei, and CrebA, a transcription factor that is expressed in the proventricular endoderm cell nuclei. These two cell types are joined at stage 12, and then undergo morphogenic movements where the ectodermal cells invaginate in towards the endodermal cells (Fuss et al 2004). By the end of development, the proventriculus is a U-shaped tube that links the foregut to the midgut. The joining of the ectodermal and endodermal cells of the proventriculus is shown in Figure 4B, and this morphology is unchanged in the *Tre1* mutants.

Over- and mis-expression of Tre1 causes a range of defects in multiple tissues

The very subtle apical membrane defects observed in the SG and other tissues in *Tre1* mutants may be due to functional redundancy with other GPCRs also in the salivary gland. To explore this possibility, we developed tools to over-express *Tre1* by placing the *Tre1* ORF downstream of Upstream Activated Elements (UAS). This method allows for tissue specific over- or mis-expression of *Tre1* using various Gal4 drivers (Brand and Perrimon, 1993). In addition to

Figure 4: Loss of *Tre1* does not affect the cuprophilic cells of the midgut or the formation of the proventriculus. A. Labial staining of the cuprophilic cells of the midgut in wild-type and *Tre1* mutants at stage 15. **B.** Fkh (red) and CrebA (green) staining of the endodermal and ectodermal portions of the proventriculus at stage 15.


WT
ΔEP19
ΔEP5
Tre1KO
αFkh αCrebA
WT
ΔEP19
ΔEP5
Tre1KO

the UAS-*Tre1* construct, we also generated a UAS-*Tre1*-GFP construct, which placed GFP in frame at the C-terminal end of Tre1. This line was used to localize *Tre1* in salivary gland cells, since our attempts at generating Tre1 antisera with both N- and C-terminal protein fragments was unsuccessful, as were the attempts of others (**Figure 5A-B''**'; Yoshiura et al., 2012). Over-expression of *Tre1* in the salivary gland by Fkh-Gal4 caused minor salivary gland defects (**Figure 5C''**), (Figure 5D). Over-expression of *Tre1*-GFP caused no major salivary gland defects (**Figure 5C'''**), suggesting that too much *Tre1* in the salivary gland does not adversely affect migration (Figure 5D). In the salivary gland, Tre1-GFP localized to the membranes, which is the expected localization of a GPCR (**Figure 5E**). Tre1-GFP colocalizes with the apical marker SAS (**Figure 5E''**) and partially overlaps with αspectrin (**Figure 5F''**), although it is more apically enriched. Therefore, Tre1-GFP is a clear marker for the lateral and apical membrane of the salivary gland, and it does not otherwise disrupt salivary gland migration.

Both constructs were also used to over-express *Tre1* in germ cells, a tissue where *Tre1* is known to function. Over-expression of untagged *Tre1* caused the germ cells to coalesce neatly into the gonad (**Figure 2H, H'**) and, in fact, had fewer germ cell mis-migrate than wild-type embryos. Tre1-GFP, however, had the opposite effect, and more closely resembled the $Tre1^{KO}$ phenotype (**Figure 2I, I'**). This suggests that the GFP tag may interfere with single-cell migration, while multicell migration (as in the salivary gland) is unaffected. Efforts to visualize Tre1-GFP in the germ cells were unsuccessful, but may need amplification methods to work, as has been shown with other germ cell proteins (Ismat et al 2013).

Using btl-Gal4, *UAS-Tre1* was mis-expressed in the trachea, a tissue that does not normally express the gene (**Figure 6A-D'**). Tracheal development in all branches of these animals was normal in early stages but defects were observed in the dorsal branches in late embryos. By embryonic stage 17, the fusion cells of almost all dorsal branches from either side of WT

Figure 5: Construction of *Tre1* tools for protein detection and over-expression analysis.

A. Topology cartoon of Tre1. The sequence in blue at the N-terminus were used for making the TreN antibody. The sequence in green at the C-terminus were used to make the TreC antibody. The red arginine indicates the conserved residue from Kamps et al (2010) which is the key residue for Tre1 function and germ cell migration. **B-B'''.** Staining with the the TreN and TreC antibodies at 1:5000. B-B'. TreN shows salivary gland staining in wild-type embryos (B). This staining is seen in $Tre1^{KO}$ embryos and is therefore unlikely to be specific. **B''-B'''**. TreC antibody staining is not observed in wild-type embryos (B'') and is not above background levels in *Tre1* over-expressing salivary glands (B'''). **C-C'''.** Over-expression of *Tre1* in the salivary gland. UAS-*Tre1* and UAS-*Tre1*-GFP constructs were made using the open reading frame of full-length *Tre1-A*. The apical membrane was stained using Crb. **D.** Quantification of apical membrane irregularities with *Tre1* over-expression in the salivary gland. **E-F''.** UAS-*Tre1*-GFP localizes to the basolateral and apical membranes but is enriched at the apical surface. Fkh-Gal4 was used to drive UAS-*Tre1*-GFP in the salivary gland. GFP staining is shown in green and SAS (E,E'') or αspectrin (F',F'') is shown in red.

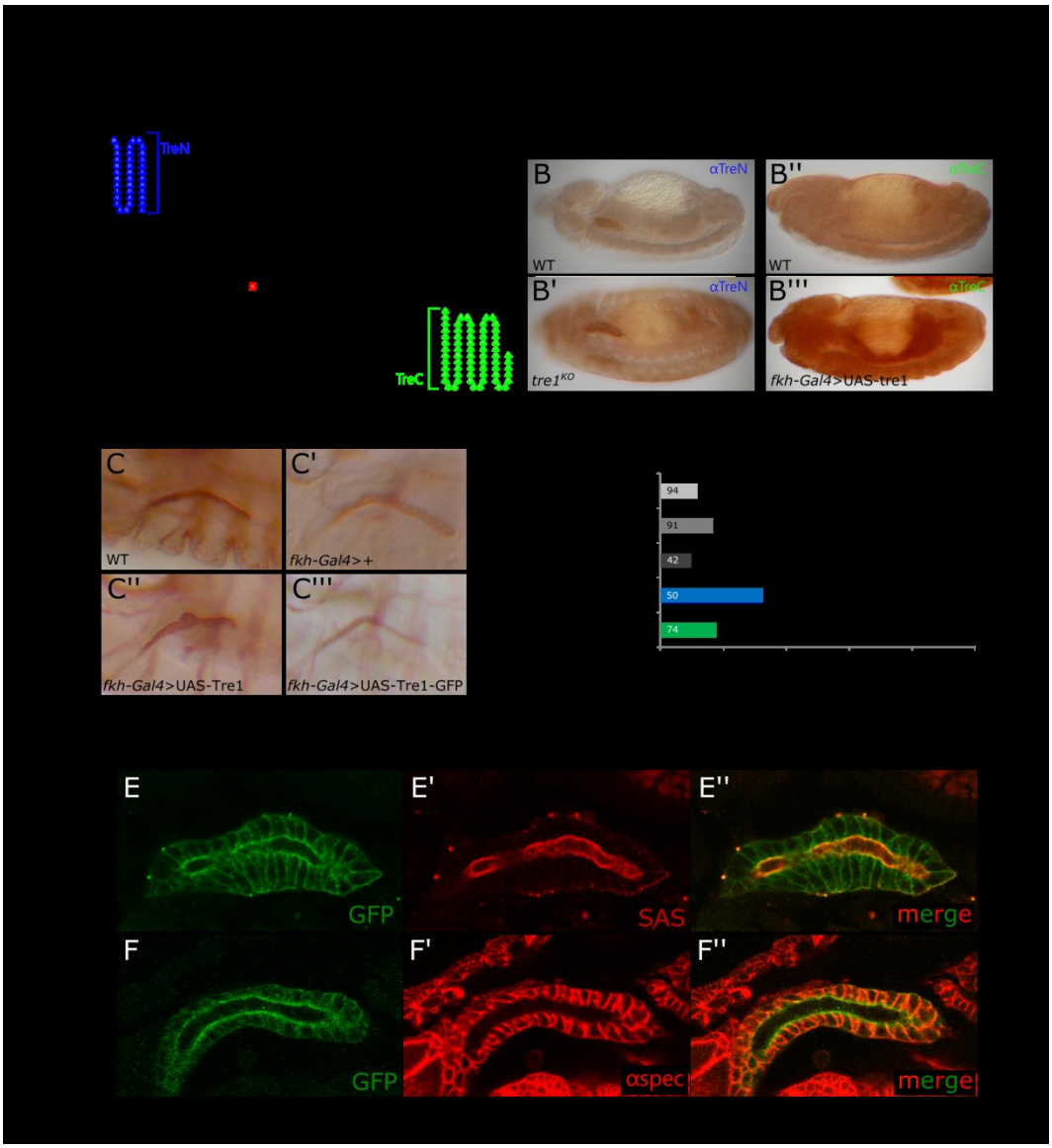
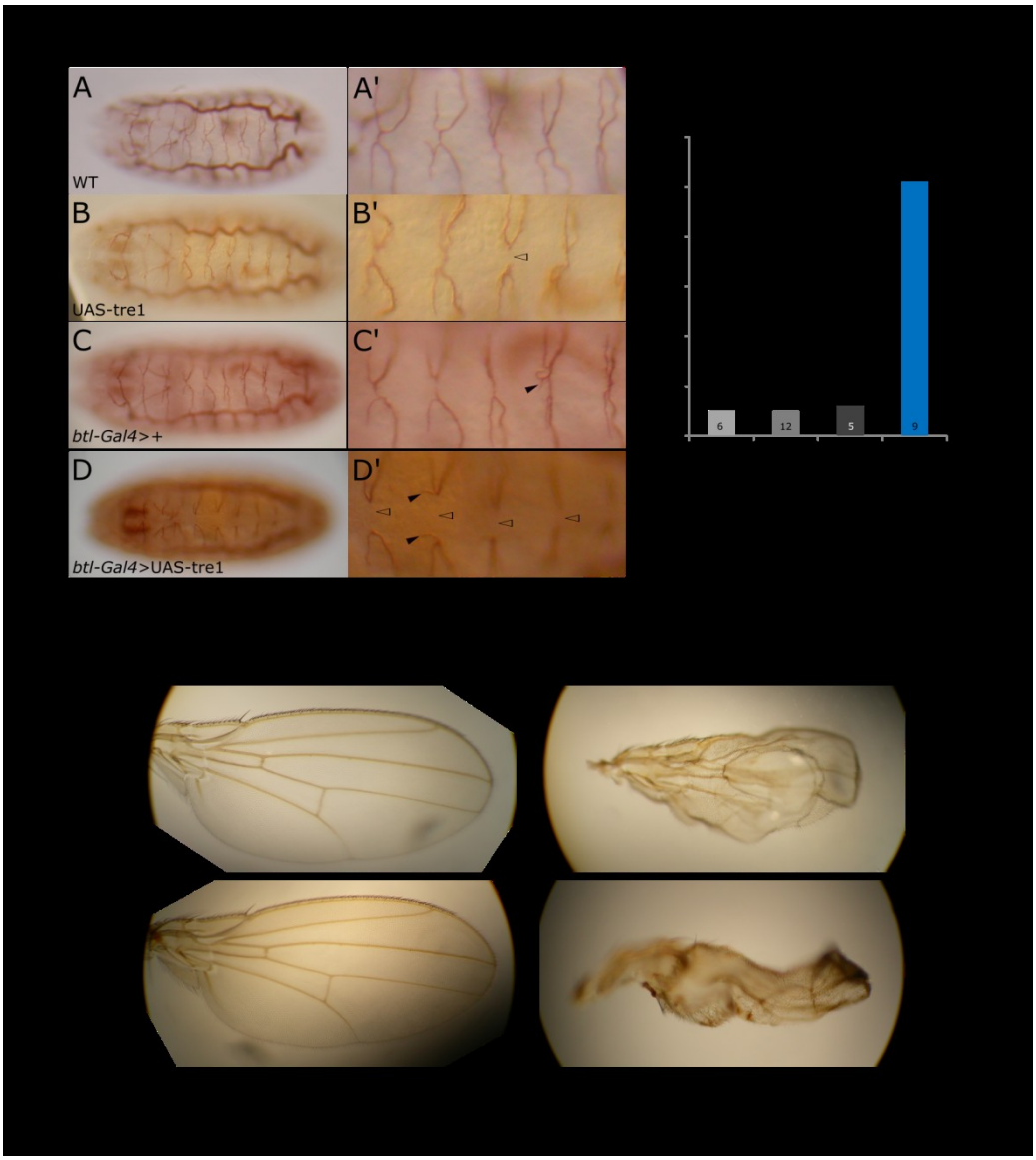

TreN
TreC
B
αTreN
WT
B''
αTreC
WT
B'
αTreN
tre1KO
B'''
αTreC
fkh-Gal4>UAS-tre1
C
WT
C'
fkh-Gal4>+
C''
fkh-Gal4>UAS-Tre1
C'''
fkh-Gal4>UAS-Tre1-GFP
94
91
42
50
74
E
GFP
E'
SAS
E''
merge
F
GFP
F'
αspec
F''
merge

Figure 6: Mis-expression of *Tre1* in the trachea and adult wing. A-D'. Mis-expression of *UAS-Tre1*in the trachea using btl-Gal4 at stage 17. The tracheal lumen is stained with 2A12. A'-D' are zoomed in to metameres three through seven. Open arrowheads show lack of fusion of the dorsal branch. Closed arrowheads show additional terminal branches that normally begin to form at this stage. **E.** Quantification of unfused dorsal branches in wild-type, control, and btl-Gal4 driven UAS-*Tre1* trachea. **F-F'.** Mis-expression of *Tre1* in the adult wing. MS1096-Gal4 is expressed to high levels in the dorsal portion of the wing disc. Mis-expression of *Tre1* results in blistered or overly folded wings.


A
A'
WT
B
B'
UAS-tre1
C
C'
btl-Gal4>+
D
D'
btl-Gal4>UAS-tre1
6
12
5
9

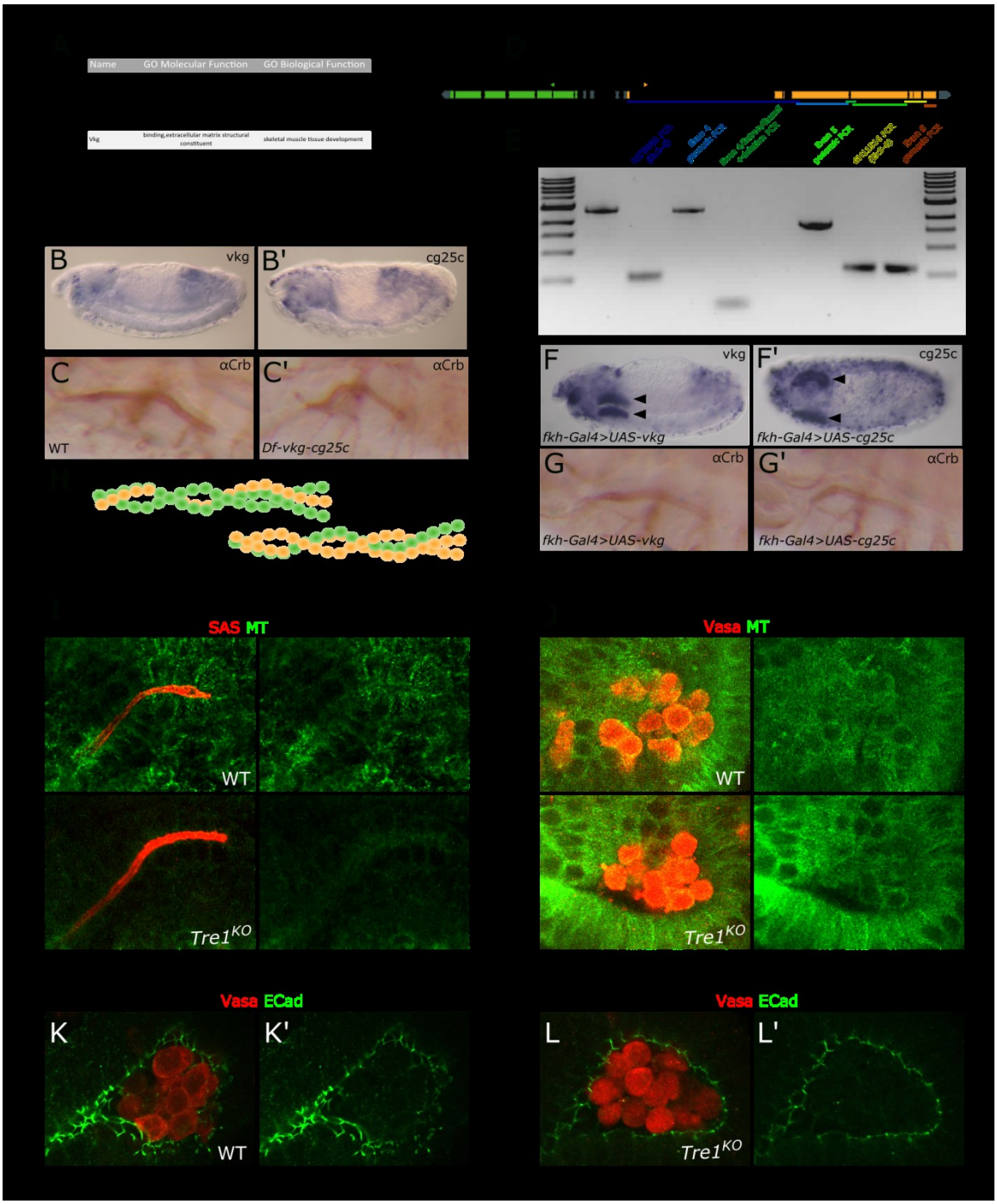
embryos have fused with their contralateral partners and the terminal tip cells have turned back toward the side of the embryo from where they originally came. Whereas a few DBs were mis-fused or unfused in both WT and control embryos (**Figure 6B',C'**), approximately half of the DBs in embryos overexpressing *Tre1* in the trachea have failed to fuse by embryonic stage 17 (**Figure 6E**).

Tre1 was also mis-expressed in the adult wing using the MS1096-Gal4 driver (**Figure 6F-F'**). MS1096-Gal4 is located on the X-chromosome, so progeny of the cross will either be heterozygous females or hemizygous males. Female wings mis-expressing *Tre1*are mis-folded and exhibit significant blistering. Male wings mis-expressing *Tre1*are severely mis-folded and do not unfurl enough to form a full wing. This suggests that *Tre1* misexpression may disrupt the normal pathways required to properly position cells during tissue morphogenesis.

Searching for a Tre1 ligand

Tre1 is an orphan GPCR and has no known ligand. To identify candidate ligands, we screened the Drosophila Protein Interactions Map database (https://interfly.med.harvard.edu/) for potential interacting molecules. This project expresses HA-tagged proteins in S2R+ cells, immunoprecipitates the proteins, then uses LC-MS to identify proteins binding to the bait molecule. The three highest hits for Tre1 binding are shown in Figure 7A. Vkg is one of the two Collagen IV molecules encoded in *Drosophila* (**Figure 7D**). Eb1 is a plus-end microtubule binding protein. We decided to investigate potential roles for these genes in salivary gland migration. Collagen IV is composed of a heterotrimer of both Vkg and Cg25c molecules (**Figure 7H**). Collagen IV is deposited by the hemocytes, cells that arise in the head and then migrate as individual cells throughout the embryo depositing basement membrane (**Figure 7B, B'**; Bunt et al 2010). A deficiency removing function of both Collagen IV genes does not affect salivary gland

Figure 7: Potential Tre1 interactors. A. Tre1 was used as a bait protein on the Drosophila Protein Interaction Map website (DPiM). The top three hits [Tre1 itself, Vkg and Eb1] are shown here, along with their Gene Ontology (GO) molecular and biological function. **B-B'.** In situ at stage 15 showing mRNA expression of the two collagen IV genes encoded in the fly. **C-C'.** Crb staining of wild-type and Df(2L)BSC172, which removes *vkg, cg25c*, and four other genes. **D.** Genomic region of *cg25c* and *vkg*. The 5' end of the *vkg* gene is closer to the centromere and the 3' end is closer to the telomere, so the cartoon is in the opposite orientation than shown in flybase. **E.** Agarose gel of the PCR fragments used to construct a *vkg* cDNA econtaining the entire coding region. Lines in D correspond to the colors of the labels in E. The names of the exons refer to their position as coding exons within the ORF. **F-F'.**Mis-expression of UAS-vkg and UAS-cg25c in the salivary gland driven by Fkh-Gal4. Arrowheads point out the salivary gland staining that is not observed in WT embryos, as shown in B and B'. **G-G'.** Crb staining of the salivary gland apical membrane in embryos mis-expressing *vkg* or *cg25c* in the salivary gland. **H.** Cartoon of collagen heterotrimer, which is composed of two Cg25c strands and one Viking strand or one Cg25c strand and two Viking strands. **I-J.** Microtubule (green) staining in the salivary gland and germ cells in wild-type and $Tre1^{KO}$ embryos. **I.** SAS (red) marks the salivary gland apical membrane. **J.** Vasa (red) marks the germ cell cytoplasm. **K-L'.** Changes in ECad localization are not observed wild-type (K-K') or $Tre1^{KO}$ (L-L') mutant embryos.


Name
GO Molecular Function
GO Biological Function
Vkg
binding,extracellular matrix structural constituent
skeletal muscle tissue development
B
vkg
B'
cg25c
C
αCrb
WT
C'
αCrb
Df-vkg-cg25c
F
vkg
fkh-Gal4>UAS-vkg
F'
cg25c
fkh-Gal4>UAS-cg25c
G
αCrb
fkh-Gal4>UAS-vkg
G'
αCrb
fkh-Gal4>UAS-cg25c
SAS MT
WT
Tre1KO
Vasa MT
WT
Tre1KO
Vasa ECad
K
K'
WT
Vasa ECad
L
L'
Tre1KO

migration (**Figure 7C,C'**), which is not surprising as hemocytes are not closely associated with the salivary gland until very late embryonic stages.

To over- or mis-express the Collagen IV molecules, UAS-vkg and UAS-cg25c constructs were made. A full-length cDNA existed for *cg25c*, but for *vkg*, a full-length cDNA had to be constructed. Using various genomic DNA fragments and available ESTs, a full-length *vkg* cDNA was made via two successive rounds of In-Fusion HD cloning. The constructs were successfully mis-expressed in the salivary gland (**Figure 7F, F'**), which did not cause apical membrane defects (**Figure G, G'**). The constructs may need to be over-expressed together so that functional heterotrimers can be formed.

To investigate the relationship between the EB1 microtubule binding protein and Tre1, we examined the microtubule organization in wild-type and $Tre1^{KO}$ embryos. EB1 binds the plus-end of microtubules, which are located at the apical membrane of the salivary gland, whereas the minus ends are at the basal membrane (Myat and Andrew, 2002). In WT salivary glands, the microtubule staining is seen at both the lateral and apical membranes, although staining was more intense near the apical surface (**Figure 7I**, top). In $Tre1^{KO}$ salivary glands, the apical staining of the MTs was more diffuse (**Figure 7I**, bottom). Microtubule staining in germ cells is comparable between wild-type and $Tre1^{KO}$ embryos (**Figure 7J**), as is germ cell morphology and E-Cad localization (**Figure 7K-L'**).

DISCUSSION

Tre1 encodes a GPCR that is expressed dynamically in tissues throughout the *Drosophila* embryo, including the germ cells, salivary gland, fat body, trunk mesoderm, proventriculus, midgut, hindgut, CNS, and pericardial cells. Mutant alleles created by p-element excision exist for *Tre1*; however, in our hands, these alleles are not null. Importantly, neither excision allele removes the start of transcription for the alternative *Tre1-B* isoform, and we observe robust *Tre1* mRNA expression from very early stages with both alleles. We therefore created a true null allele of *Tre1,* replacing entire the *Tre1* coding region, by homologous recombination. This allele is RNA null and homozygous embryos from homozygous mothers have severe germ cell migration defects than observed with homozygous embryos from mothers carrying either excision allele.

Because *Tre1* is expressed just before and during active migration of the salivary gland, we asked if loss of *Tre1* affects gland migration as it does germ cell migration. Overall salivary morphology and placement were normal in *Tre1* null embryos. The only phenotype we observed were minor apical membrane irregularities, which were present in only a small fraction of mutant embryos. Indeed, the only consistent phenotype we observed was the irregular patterns of MT and SAS apical staining. The absence of overt salivary gland migration phenotypes in the $Tre1^{KO}$ that we expected based on the severe germ cell migration defects may reflect differences between single cell and multi-cell migration. Whereas the main facets of single and multi-cell migration remain the same – sending out protrusions, integrin-mediated attachment to a substrate, contraction of the cell body, and forward movement by retraction of the lagging end – a key difference is the junctional attachments that exist between the cells in multi-cell migration (Ilina and Friedl, 2009). The salivary gland is not only a multi-cellular tissue, but it also migrates as a fully polarized epithelium, a markedly different structure from a cell cluster. Thus,

if most of the cells "decide" to move in a given direction, the entire tissue is likely to move. If so, the irregularities observed at the apical surface in the *Tre1* mutants reflect some non-consensus in these migration decisions. Importantly, the SG receives input from multiple different tissues, including the visceral and somatic mesoderm, the fat body, and the CNS, in order to migrate to the correct position. It remains unclear when and how many signals are made by these surrounding tissues, so removing only a single signal may not be expected to have major consequences (Vining et al 2005; Harris and Beckendorf, 2007, Kolesnikov and Beckendorf, 2005).

Another difference between germ cell migration and salivary gland migration is in the number of cells it takes to create a functional tissue. The salivary gland is composed of about 140 cells per placode, and as the gland invaginates and migrates, there is no cell death or division. Moreover, in embryos with fewer cells in the salivary gland placode, the gland fails to form correctly (Lammel et al 2000). This is in contrast to germ cell migration, where if even a single pole cell reaches the gonad as a primordial germ cell, that cell will proliferate to fill the germ cell niche. Thus, the gonad forms correctly and the organism is fertile (Robertson et al 1999). In this way, collective cell migration is more of an all-or-nothing endeavor, which is potentially why there appear to be so many inputs and redundancies in the salivary gland migration system. Indeed, one reason why salivary gland migration is unaltered in $Tre1^{KO}$ may be because of redundancy with another salivary gland expressed GPCR, *mthl5*, or redundancy with other signaling pathways such as Slit or Robo (Kolesnikov and Beckendorf, 2005). Such possibilities can now be explored with the null *Tre1* allele that we have generated.

An alternative hypothesis is that *tre1* does not function in SG migration at all and is instead required for the organization of apically localized proteins, such as Ebi1. This would be consistent with the reported roles for Tre1 in both germ cells and in neuroblasts, and with the

high levels of Tre1-GFP accumulation at the apical surface. Recently, a genetic interaction was found between *Tre1* and *tao-L*, a microtubule destabilizing protein (Pflanz et al 2015). While no direct mechanism has been described, the authors hypothesize that activated Tre1 activates the Tao-L kinase, which causes the formation of lammelipodia-like structures. Tao-L is not expressed in the salivary gland, but this finding nonetheless supports the idea of Tre1 being involved in cytoskeletal regulation of migrating tissues. The method that was used to identify Ebi1 as a Tre1 interactor also identified Tre1 itself. This suggests that Tre1 oligamerizes with itself, which may aid in concentrating the receptor apically (Gurevich and Gurevich, 2008).

Tre1 is thought to affect in germ cell migration by changing the localization of Rho1, Gβ13f, and E-Cadherin, which alters cell polarity and adhesion to enable migration across the endoderm (Kunwar et al 2008). Whereas these observations explain the trapped germ cell phenotype, it does not easily explain why germ cells are subsequently found scattered throughout the embryo. Importantly, we did not observe the same changes in E-Cadherin localization in $Tre1^{KO}$ embryos that had been reported with the excision alleles, but this could simply be because E-Cadherin is a finicky antibody we may have missed the time window when those differences can be observed. If E-Cadherin relocalization is delayed or altered and the germ cells cross the endoderm later than normal, they may miss interactions with the migrating somatic gonadal precursors (SGP). This might explain why germ cells are scattered in the embryo. Another possibility is that *Tre1* plays a separate role in germ cell guidance once the germ cells are through the midgut and this second function may not be disrupted in the excision alleles.

Several factors affect germ cell migration after exit from the midgut. Wunen and Wunen2 are lipid phosphate phophohydrolases that dephosphorylate extracellular lipid phosphates such as phosphatidic acid and lysophosphatidic acid. Germ cells, through an unknown mechanism,

migrate towards phosphorylated lipid phosphates, so the expression of *wun* and *wun2* in the rear of the germ cells and in the soma drives germ cells away (Renault, et al 2004; Starz-Gaiano et al 2001). The HMG-CoA reductase (Hmgcr) pathway also contributes to germ cell migration. *hmgcr* is expressed in the somatic gonadal precursors (SGP), the destination for the germ cells after traversing the endoderm. The Hmgcr pathway results in the farnesylated or geranylgeranylated proteins (Ding et al 2008). One theory is that Hmgcr modifies a protein, which is then secreted through an ABC transporter to attract the germ cells (Ricardo and Lehmann 2009; Santos and Lehmann 2004). Another theory is that Hmgcr targets Gγ1, which promotes hedgehog (Hh) trafficking, and that Hh is the germ cell attractant made in the SGP (Deshpande et al 2005; Deshpande et al 2009). Both possibilities could be true, but transporter-Hh mutants have not yet been investigated for a genetic interaction.

Ectopic expression of *Tre1* in both the embryonic trachea and the wing disc result in defects that are consistent with cell moving in the wrong direction, perhaps in the direction of a source for the unknown Tre1 ligand. Alternatively, ectopic expression of Tre1 may have dominant negative effect in these tissues. In the wing, over-expression of Gαo inhibits the activity of Gαs by sequestering Gβ13f and Gγ1 (Katayaneva et al 2010). Tre1 activation could do something similar. In any case, finding the endogenous ligand for the widely expressed Tre1 protein would be an enlightening development in studies of a GPCR that is expressed in so many embryonic cell types.

REFERENCES

Andrew D.J., Baig A., Bhanot P., Smolik S.M., and Henderson K.D. (1997). The *Drosophila* dCREB-A gene is required for dorsal/ventral patterning of the larval cuticle. Development. **124**:181–193

Andrew, D.J., Horner, M.A., Petitt, M.G., Smolik, S.M., and Scott, M.P. (1994). Setting limits on homeotic gene function: restraint of Sex combs reduced activity by teashirt and other homeotic genes. *EMBO J.* **13**(5): 1132-1144.

Bradley, P.L., Myat, M.M., Comeaux, C.A., and Andrew, D.J. 2003. Posterior migration of the salivary gland requires an intact visceral mesoderm and integrin function. *Developmental Biology* **257**: 249-262.

Brand, A.H., Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**(2): 401--415.

Bunt, S., Hooley, C., Hu, N., Scahill, C., Weavers, H., Skaer, H. (2010). Hemocyte-Secreted Type IV Collagen Enhances BMP Signaling to Guide Renal Tubule Morphogenesis in Drosophila. *Dev. Cell* **19**(2): 296--306.

Deshpande, G., Godishala, A., Schedl, P. (2009). Ggamma1, a downstream target for the hmgcr-isoprenoid biosynthetic pathway, is required for releasing the Hedgehog ligand and directing germ cell migration. *PLoS Genet*ics **5**(1): e1000333.

Deshpande, G., and Schedl, P. (2005). HMGCoA reductase potentiates hedgehog signaling in Drosophila melanogaster. *Dev. Cell* **9**(5): 629--638.

Ding, J., Jiang, D., Kurczy, M., Nalepka, J., Dudley, B., Merkel, E.I., Porter, F.D., Ewing, A.G., Winograd, N., Burgess, J., and Molyneaux, K. (2008). Inhibition of HMG CoA reductase reveals an unexpected role for cholesterol during PGC migration in the mouse. *BMC Developmental Biology*, **8**:120**.**

Doitsidou, M., Reichman-Fried, M., Stebler, J., Koprunner, M., Dorries, J., Meyer, D. Esquerra, C.V., Leung, T., and Raz, E. (2002). Guidance of primordial germ cell migration by the chemokine SDF-1. *Cell*, **111**, 647-59.

Fox, R.M., Vaishnavi, A., Maruyama, R., and Andrew, D.J. (2013). Organ-specific gene expression: the bHLH protein Sage provides tissue specificity to Drosophila FoxA. *Development* **140**(10): 2160-2171.

Gong, W.J., and Golic, K.G. (2003). Ends-out, or replacement, gene targeting in Drosophila. *PNAS* **100**(5): 2556-2561.

Gurevich, V.V., and Gurevich, E.V. (2008). How and why do GPCRs dimerize? *Trends Pharmacol Sci* **29**(5): 234-240.

Harris, K.E. and Beckendorf, S.K. 2007. Different Wnt signals act through the Frizzled and RYK receptors during Drosophila salivary gland migration. *Development* **134**: 2017-2025.

Henderson, K.D., and Andrew, D.J. (2000). Regulation and function of Scr, exd, and hth in the Drosophila salivary gland. *Dev. Biol.* **217**(2): 362-374.

Ismat, A., Cheshire, A.M., and Andrew, D.J. (2013). The secreted AdamTS-A metalloprotease is required for collective cell migration. Development 140(9): 1981--1993.

Kamps, A.R., Pruitt, M.M., Herriges, J.C., and Coffman, C.R. (2010). An Evolutionarily Conserved Arginine Is Essential for Tre1 G Protein-Coupled Receptor Function During Germ Cell Migration in Drosophila melanogaster. *PLoS ONE* **5**(7): e11839.

Katanayeva, N., Kopein, D., Portmann, R., Hess, D., and Katanaev, V. L. (2010). Competing activities of heterotrimeric G proteins in Drosophila wing maturation. *PloS One*, **5**, e12331.

Kolesnikov, T. and Beckendorf S.K. (2005). Netrin and Slit guide salivary gland migration. *Developmental Biology* **284**(1): 102-111.

Kunwar, PS, Sano, H, Renault, AD, Barbosa, V, Fuse, Naoyiku, and Lehmann, R. (2008). Tre1 GPCR initiates germ cell transepithelial migration by regulating *Drosophila melanogaster* E-Cadherin. *JCB* **183**(1): 157-168.

Kunwar, PS, Starz-Gaiano, M, Bainton, RJ, Heberlein, U, and Lehmann, R. (2003). Tre1, a G-protein coupled receptor, directs transepithelial migration of *Drosophila* germ cells. *PLOS Biology* **1**(3): 372-384.

Lammel, U., Meadows, L., and Saumweber, H. (2000). Analysis of Drosophila salivary gland, epidermis and CNS development suggests an additional function of brinker in anterior-posterior cell fate specification. *Mech. Dev.* **92**(2): 179-191.

Lehmann, R., and Tautz, D. (1994). In situ hybridization to RNA. *Mehtods Cell Biol* **44**: 575-98.

Molyneaux, K. A. (2003). The chemokine SDF1/CXCL12 and its receptor CXCR4 regulate mouse germ cell migration and survival. *Development*, **130**, 4279–4286.

Myat, M.M., and Andrew, D.J. (2002). Epithelial tube morphology is determined by the polarized growth and delivery of apical membrane. *Cell* **111**(6): 879-891.

Pflanz, R., Voigt, A., Yokulov, T., and Jackle, H. (2015). Drosophila gene tao-1 encodes proteins with and without a Ste20 kinase domain that affect cytoskeletal architecture and cell migration differently. *Open Biology*.

Prabhu, Y., and Eichinger, L. (2006). The Dictyostelium repertoire of seven transmembrane domain receptors. *European Journal of Cell Biology*, **85**, 937–46.

Renault, A.D., Sigal, Y.J., Morris, A.J., and Lehmann, R. (2004). Soma-germ line competition for lipid phosphate uptake regulates germ cell migration and survival. *Science* **305**(5692): 1963-1966.

Reuter, R., and Scott, M.P. (1990). Expression and function of the homoeotic genes Antennapedia and Sex combs reduced in the embryonic midgut of Drosophila. *Development* **109**:289–303.

Ricardo, S., and Lehmann, R. (2009). An ABC transporter controls export of a Drosophila germ cell attractant. *Science* **323**(5916): 943-946.

Robertson, S.E., Dockendorff, T.C., Leatherman, J.L., Faulkner, D.L., and Jongens, T.A. (1999). germ cell-less is required only during the establishment of the germ cell lineage of Drosophila and has activities which are dependent and independent of its localization to the nuclear envelope. *Dev. Biol.* **215**(2): 288-297.

Santos, A.C., and Lehmann, R. (2004). Isoprenoids control germ cell migration downstream of HMGCoA reductase. *Dev. Cell* **6**(2): 283-293.

Shiga, Y., Tanaka-Matakatsu, M., and Hayashi, S. (1996). A nuclear GFP/ beta-galactosidase fusion protein as a marker for morphogenesis in living Drosophila. *Dev. Growth Differ.* **38**(1): 99--106.

Starz-Gaiano, M., Cho, N.K., Forbes, A., and Lehmann, R. (2001). Spatially restricted activity of a Drosophila lipid phosphatase guides migrating germ cells. *Development* **128**(6): 983--991.

Tremml, G., and Bienz, M. (1992). Induction of labial expression in the Drosophila endoderm: response elements for dpp signalling and for autoregulation. *Development* **116**(2): 447-456.

Vining, M.S., Bradley, P.L., Comeaux, C.A., and Andrew, D.J. (2005). Organ positioning in Drosophila requires complex tissue-tissue interactions. *Dev Biol*. **287**(1):19-34.

Waller-Evans, H., Promel, S., Langenham, T., Dixon, J., Zahn, D., Colledge, W.H., Doran, J., Carlton, M.B.L., Davies, B., Aparicio, S.A.J.R., Grosse, J., and Russ, A.P. (2010). The

orphan adhesion GPCR GPR126 is required for embryonic development in the mouse. *PLoS One*, **5**(11), e14047.

Wang, H., Ng, K.H., Qian, H., Siderovski, D.P., Chia, W., and Yu, F. (2005). Ric-8 controls Drosophila neural progenitor asymmetric division by regulating heterotrimeric G proteins. *Nat. Cell Biol.* **7**(11): 1091--1098.

Yoshiura, S., Ohta, N., and Matsuzaki, F. (2012). Tre1 GPCR signaling orients stem cell divisions in the Drosophila central nervous system. *Developmental Cell*, **22**, 79–91.

CHAPTER 4:

The Role of the GPCR *mthl5* in Salivary Gland Invagination

ABSTRACT

The Methuselah family is the third largest family of G-protein coupled receptors (GPCRs) in *Drosophila*, yet most members of this family are uncharacterized. Recently, *mist* (*mthl1)* was identified as the receptor for *fog*, which causes activation of the Rho signaling pathway that ultimately results in the cell shape changes that drive gastrulation. *methuselah like 5* (*mthl5)* is transiently expressed in the salivary gland during tissue invagination. The initial trigger for the apical constriction that precedes gland invagination remains unknown. Because *mthl5* and *mist* share a high level of sequence identity, we investigated if *mthl5* can direct the cell shape changes that accompany internalization of the salivary gland. We created *mthl5* null alleles using CRISPR/Cas9 and as well as constructs for overexpression of *mthl5*. We found that salivary gland cells lacking *mthl5* do not decrease the size of their apical membrane prior to and during invagination but nonetheless invaginate to form internalized salivary glands with abnormally shaped lumens. Moreover, E-Cadherin, which is known to play a role in cell invagination (Wang et al 2013), is more diffuse in *mthl5* null flies. Staining of glands with Mthl5 antiserum reveals that the protein localizes to the apical surface of salivary gland cells during invagination, but that at later stages of gland migration, Mthl5 localizes to the basolateral membranes. We next questioned if Mthl5 and Mist share a ligand. *Fog* is expressed in the salivary gland at the same time as *mthl5.* Both knock-down and over-expression of *fog* cause defects in the salivary gland apical membrane, and related defects are seen with knock-down and over-expression of *mthl5*. Altogether, these data suggest that Fog acts through Mthl5 to drive salivary gland invagination. Currently, we are investigating if Fog and Mthl5 activate the Rho signaling pathway in the salivary gland, which would provide a mechanism for the initial steps of salivary gland invagination.

Introduction

G-protein coupled receptors (GPCRs) are among the largest classes of protein families in all higher organisms (Kobilka 2007). All GPCRs share a seven-transmembrane architecture, allowing them to interact with the extracellular environment and transmit pertinent information to the inside of the cell. Despite a common membrane organization, GPCRs are quite different from one another in amino acid sequence. These differences allow GPCRs to sense astoundingly diverse ligands, from light to hormones to peptides. When a GPCR binds a ligand, it undergoes a conformational change that is transduced to the inside of the cell, resulting in activation of the heterotrimeric G-proteins (Gα, Gβ, and Gγ). In turn, G-proteins activate second messenger cascades that alter cell behavior. Unsurprisingly, aberrant activation or silencing of GPCR signaling can result in deleterious cellular behavior, contributing to diseases such as cancer or deaf-blindness (O'Hayre et al, 2014). The involvement of GPCR signaling in such a wide range of diseases has made the family a focus of intense biomedical research, with nearly 40% of all drug development targeting GPCRs.

The GPCR superfamily is subdivided into subfamilies based on sequence similarity. One such family is the Secretin family, whose members are characterized by their long extracellular N-terminus (Yona et al 2008). The N-terminus is believed to assist in the recognition and binding of peptide and protein hormones. Insects encode a unique subfamily of Secretin GPCRs not found in humans or worms (Harmar 2001), known as the Methuselah (Mth) family, which is largely uncharacterized. Sixteen Mth family members are encoded in the Drosophila genome (**Figure 1A**) but only two of these family members have been characterized to any degree. Mth, the founding member of the family, was identified in a screen for mutations that increased adult lifespan (Lin et al 1998). Through a heterologous expression system, Mth was subsequently found to bind Stunted (Sun), a subunit of the F_1-F_0-ATP synthase, and activate calcium signaling

(Cvejic et al 2004). Supporting this interaction, mutations in Sun also increase lifespan in adult flies. Sex Peptide (SP) was later identified as a Mth ligand in HEK293 cells, but it remains unclear if the interaction holds in vivo (Ja et al 2009).

The other member of the Drosophila Mth/Secretin family that has been characterized is Mist, or Methuselah-like 1 (Mthl1; Manning et al 2013). As opposed to the Mth studies, the receptor was identified long after the ligand. The ligand Folded Gastrulation (Fog) had been identified as one of the key molecules driving embryonic gastrulation (Costa et al 1994). Loss of *fog* results in defects in gastrulation, a phenotype attributed to a failure of mesodermal cells to constrict apically. Other components of the Fog signaling pathway were identified by either shared phenotypes or genetic interactions. They include the Gα subunit concertina (cta; Parks and Wieschaus, 1991) and the G-protein related kinase 2 (Gprk2; Fuse et al 2013), implicating a GPCR as a key mediator for gastrulation. Mist was identified through an RNAi cell culture screen as the receptor for Fog (Manning et al 2013). Importantly, this interaction holds in vivo; mutations in *mist* also cause mild gastrulation defects. With its known ligand, downstream Gα protein and a regulating kinase, Mist is the best-characterized Drosophila Secretin family member.

Whereas gastrulation is one of the earliest coordinated cell invagination event in the Drosophila embryo, many other cells and tissue undergo invagination later in development. One such organ is the salivary gland (Myat and Andrew, 2000a). The salivary gland begins as a placode of polarized columnar epithelial cells on the surface of the embryo that invaginates and moves dorsally until it contacts the visceral mesoderm. There, it reorients and begins posterior migration, continuing until the gland reaches its final position along the anterior-posterior axis of the embryo (Vining et al 2005). Much like in gastrulation, salivary gland invagination involves cell shape changes (Myat and Andrew, 2000b). After placode formation, cells in the dorsal-

posterior quandrant of the placode begin to apically constrict and internalize. This constriction is linked to changes in Myosin activity, and mutations in myosin upstream activators like Rho and Rho kinase have salivary gland invagination defects (Roper 2012; Xu et al 2008). Salivary gland invagination completely fails in embryos with loss-of-function mutations in the gene encoding the winged helix transcription factor Fork head (Fkh). In *fkh* mutant embryos, the salivary glands remain as placodes on the embryo surface (Myat and Andrew, 2000a). The molecules downstream of Fkh that drive salivary gland invagination have yet to be characterized.

Since the known receptor that drives invagination of the mesoderm – Mist – is not expressed in the salivary gland, it is likely that another member of the Mth family could function as a GPCR for salivary gland internalization. An excellent candidate GPCR is methuselah-like 5 (*mthl5*), which is expressed early during salivary gland invagination and disappears during gland migration (Fig 1C). Mthl5 is most closely related to the Mth family members Mthl15, Mthl14, and Mist. Beyond a reported expression pattern, very little is known about *mthl5* (Patel et al, 2012). Based on its expression pattern and its homology to Mist, a family member required for mesodermal invagination, we decided to characterize *mthl5* and learn its role in salivary gland development.

MATERIALS AND METHODS

Fly Strains

The mthl5 CRISPR sites were chosen using the CRISPR finder developed by the Perrimon lab (http://www.flyrnai.org/crispr/). Oligos were ligated into the pU6-BSA-gRNA plasmid (Baena-Lopez et al, 2013) after BsaI digestion. Constructs were injected into lines y[1] P{vas-Cas9.S}ZH-2A w[1118] flies by Rainbow Transgenic Flies, Inc (Camarillo, CA). Cloning of the oligos into the appropriate vector and the initial fly crosses were performed by Arun Sridharan, an undergraduate in the lab. UAS-*mthl5* was generated by Gateway cloning (Invitrogen, Carlsbad, CA) of the complete *mthl5* open reading frame into the pTW UAS vector. This construct was then injected into w^{1118} recipient flies by Rainbow Transgenic Flies, Inc (Camarillo, CA) and transgenic animals were identified by the $white^{+}$ eye color marker. *Df(3R)BSC514*, *Df(3R)Exel7310*, and fog^{s4} were obtained from the Bloomington Stock Center (Bloomington, Indiana). UAS-*fog* was a gift from the Perrimon Lab (Hacker et al 1998).). *twi*-Gal4 was used to drive UAS-mthl5 in the mesoderm (Greig and Akam, 1993). *fog* and *mthl5* RNAi lines were obtained from the Vienna Drosophila Resource Center (VDRC; Vienna, Italy). *fkh*-Gal4 was created by our laboratory (Henderson and Andrew, 2000). *en*-Gal4 (Weiss et al., 2001) was used to express UAS-fkh (Maruyama et al., 2011) in the ectodermal stripes.

Mthl5 antibody generation

Genomic fragments of the first 648 base pairs (corresponding to N-terminal extracellular 216 amino acids) were cloned in-frame into the pET15b expression vector (Novagen) using the In-Fusion HD Cloning Kit (Clontech). The construct was transformed into *E Coli*. Protein was induced during log growth phase with 0.1M IPTG and grown over-night at 16^{o}C. Inclusion body

preps were performed to isolate the induced protein, which was subsequently used to inoculate guinea pigs (Covance).

Immunohistochemistry and in situ hybridizations

In situ hybridization and immunohistochemistry were performed as previously described (Lehmann and Tautz, 1994; Reuter et al., 1990). *mist* and *mthl14 in situ* probes were made by cloning genomic fragments into pCR-TOPO-II vector (Invitrogen). Antibody concentrations used in this study are as follows: GPαMthl5 (1:500), RbαFork head (1:2000; a gift from S. Beckendorf, Berkeley, CA, USA), RtαCrebA (1:1,000; Andrew et al., 1997), GPαSage (1:500; Fox et al 2013), RbαCrebA (1:10,000; Fox et al., 2010), RbαSAS (1:500; a gift from D. Cavener, Penn State University, PA, USA), RbαGFP (1:500, Invitrogen), mααSpectrin (1:2; Developmental Studies Hybridoma Bank, Iowa City, IA), RbαNrxIV (1:5000; H. Bellen, Baylor College of Medicine, Houston, TX, USA), mαCrumbs (1:100, DSHB), RtαDE-Cad (1:10, DSHB), and mαβgalactosidase (Promega, 1:10,000). All secondary antibodies (Vector Labs and Molecular Probes) were used at a 1:500 dilution. HRP images were developed using the Vectastain ABC kit (Vector Laboratories). Confocal images were obtained using a Zeiss LSM 510 or 700 Meta confocal microscope, using a Plan-Neofluor 63x and 100x, 1.3 oil objective and the Zeiss LSM software. All other images were obtained using a Zeiss Axiophot microscope configured with a Nikon Coolpix 4500 digital camera or a Janoptik ProgResC14 Plus opitical imaging system. Images were taken using a Plan-Neofluor 20x, 0.50 objective. Images were rotated and cropped using Inkscape, an open-source vector graphic editor. All images were obtained at room temperature.

RESULTS

***mthl5* is transiently expressed in the early salivary gland**

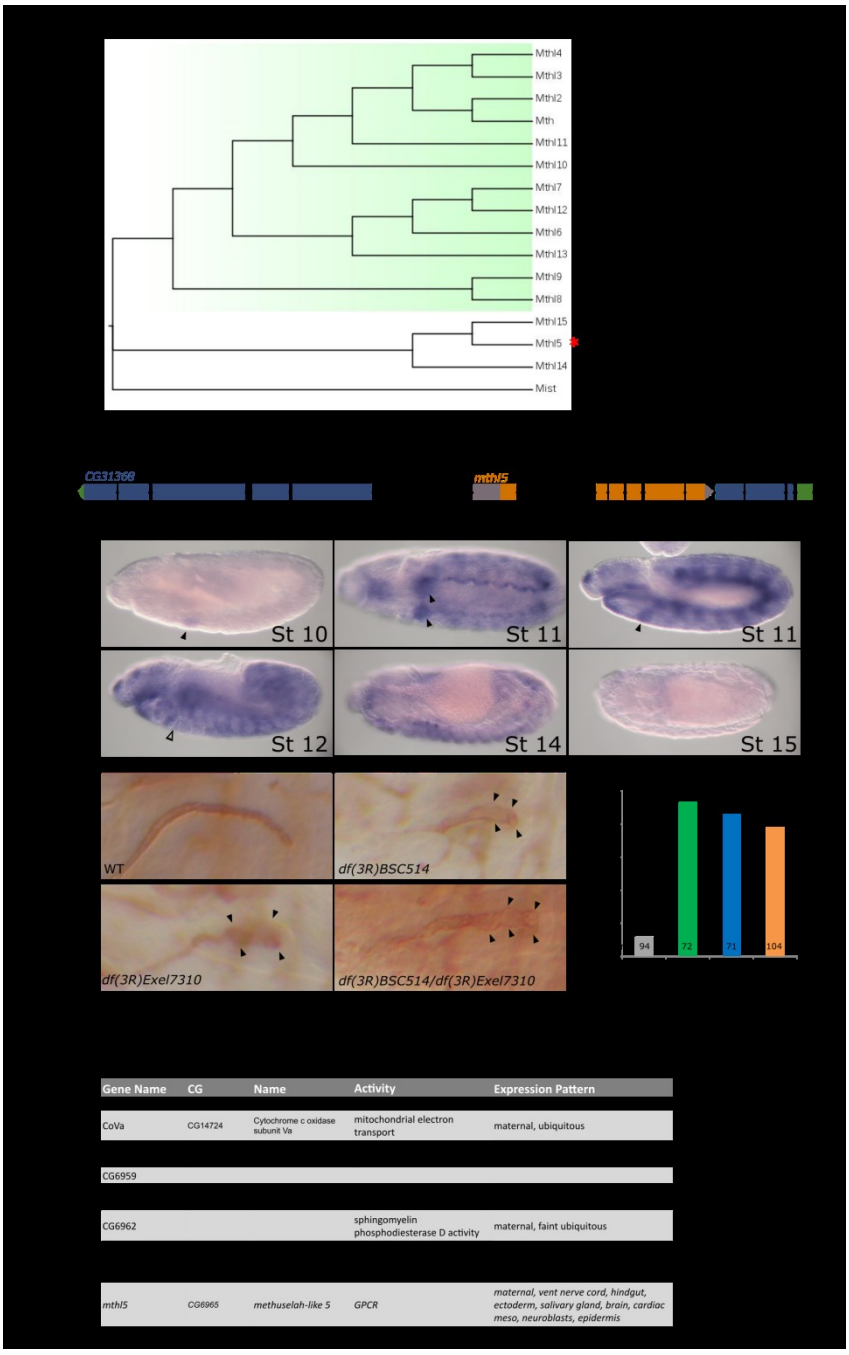
Sixteen *mth* family members are encoded in Drosophila. Protein sequences corresponding to these genes were used to generate a rooted phylogenetic tree based on Clustal W2 alignments (**Figure 1A**). Twelve family members separated into the previous described Mth superclade, with Mthl14, Mthl5, Mthl15, and Mist separating into different clades. Notably, Mist separates into a distinct clade on its own, Mthl14, Mthl15 and Mthl5 are in another, with Mthl5 and Mthl15 segregating together into an even closer grouping.

The *mthl5* gene maps to the right arm of chromosome 3 and is entirely contained within an intron of the uncharacterized *CG31368* gene (**Figure 1B**). *mthl5* has six exons, with the start codon in the first exon and the stop codon in exon 6. *mthl5* is expressed in a dynamic pattern during embryonic development (**Figure 1C**). *mthl5* transcript is detected in several invaginating tissues, including the salivary gland (arrowheads), mesodermal crest cells, and trachea. In the salivary gland, *mthl5* is expressed prior to (st 10) and during salivary gland invagination (late st 11, early st 12), and disappears when active migration begins (late st 12 and later). *mthl5* expression is also detected in the wing and leg haltere imaginal discs (st 14), as well as in regions in the head.

Loss of mthl5 disrupts salivary gland morphology

Because of the early salivary gland expression of *mthl5*, we asked if its loss affects salivary gland morphology. Two deficiency lines that remove *mthl5* were obtained. *Df(3R)BSC514* removes 48 genes and *Df(3R)Exel7310* removes 13 genes. When in trans to one another, the two deficiencies remove only eight genes, including *mthl5* (**Figure 1E**). By stage 14, the wild-type salivary gland is fully internalized and is positioned along the anterior-posterior axis. The apical

Figure 1: *mthl5* is expressed in the early stages of salivary gland formation. A. Mthl5 (red asterisk) is a member of the Methuselah (Mth) subfamily of Secretin GPCRs in *Drosophila*. Mthl15, Mthl5, Mthl14, and Mist are separate from the main Mth family superclade (green). **B**. *mthl5* is encoded on the third chrosomome, completely with the fourth intron of *CG31368*. *mthl5* transcript is indicated by the gray box; coding exons are indicated by the orange boxes. For CG31368, transcript is indicated by the green boxes; coding exons are indicated by the blue boxes. *mthl5* is encoded on the antisense strand, so this depiction is reversed from the orientation in the genome.**C.** *mthl5* is expressed dynamically during embryogenesis. Lateral views are shown for each developmental stage, except for Stage 11, which is a ventral-lateral view. During Stage 10, *mthl5* is expressed in the nascent salivary gland placode (arrowhead). By Stage 11, the expression has expanded to include mesodermal crest cells, the trachea and head. By late stage 12, expression in the salivary gland is no longer detected (open arrowhead). At Stage 14, expression is largely down-regulated in the embryo, except in the leg and wing imaginal disk primordia. By Stage 15, *mthl5* expression is not detected in the embryo. **D-D'.** Deficiencies removing *mthl5* show defects in the salivary gland apical membrane. Crb staining in wild-type embryos is regular and smooth, but is often inflated and irregular in the deficiency lines. Quantification of the irregularities is shown in (D'). **E.** *Df(3R)514* and *Df(3R)Exel7310* in trans to one another remove *mthl5* and the seven other genes shown in the table.



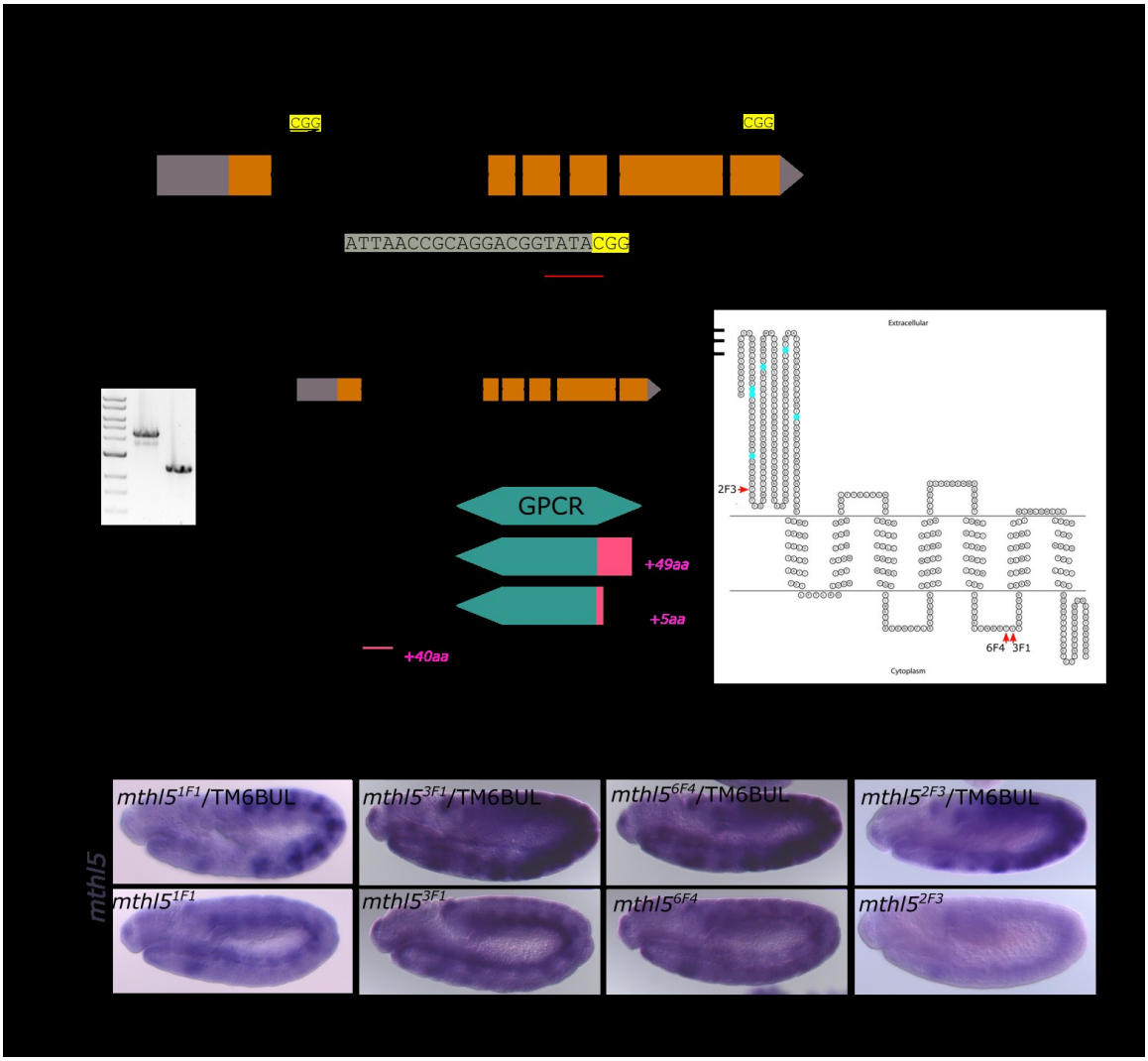
Gene Name	CG	Name	Activity	Expression Pattern
CoVa	CG14724	Cytochrome c oxidase subunit Va	mitochondrial electron transport	maternal, ubiquitous
CG6959				
CG6962			sphingomyelin phosphodiesterase D activity	maternal, faint ubiquitous
mthl5	*CG6965*	*methuselah-like 5*	*GPCR*	*maternal, vent nerve cord, hindgut, ectoderm, salivary gland, brain, cardiac meso, neuroblasts, epidermis*

membrane, as revealed by staining with the apical marker Crumbs (Crb), is smooth and has an even width (**Figure 1D**). In embryos homozygous for either deficiency or with the two deficiencies in trans to one another, the apical lumens of the salivary glands are inflated and the apical membrane is quite irregular (**Figure 1D, E**). These phenotypes are observed in over 80% of salivary glands from deficiency embryos, consistent with *mthl5* playing a role in salivary gland morphogenesis.

Creation of a *mthl5* null allele

To create a clean null allele of *mthl5*, the CRISPR/Cas9 system was used. Two CRISPR constructs targeting *mthl5* were injected simultaneously, one targeting the first coding exon, one targeting the fifth coding exon (**Figure 2A**), with the goal of creating a large deficiency. Of the three individual mutant lines generated, two had small lesions near the Exon 5 targeted site (Fig 2B). Line 6F4 was a seven bp deletion just upstream of the PAM site, whereas line 3F1 was a deletion of three base pairs plus an insertion of four base pairs. PCR amplification of the region immediately flanking Exon 5 of the 2F3 allele repeatedly failed, so a large genomic region of *mthl5*^{2F3} was amplified instead (**Figure 2C**). This PCR amplification product was approximately two kb shorter than that generated with wild-type genomic DNA, suggesting that a large deletion had occurred. Sequencing the entire gene revealed that in the *mthl5*^{2F3} allele, the region between the first exon-intron boundary and the targeted Exon five PAM site was deleted (**Figure 2D**). The altered ORFs for the *mthl5*^{3F1}and *mthl5*^{6F4} encode a large N-terminal portion of Mthl5, including the first five of the seven transmembrane domains, before going out of frame (pink) and truncating (**Figure 2D', E**). The *mthl5*^{2F3} allele encodes only the 44 most N-terminal residues of Mthl5 plus 40 out of frame residues. Thus, *mthl5*^{2F3} is likely to be null (**Figure 2D', E**).

Figure 2: Creation of *mthl5* mutant alleles via CRISPR/Cas9. A. *mthl5* was targeted by two CRISPR constructs. The first construct targets sequences in the first coding exon; the second construct targets sequences in the fifth coding exon. PAM sites are highlighted in yellow. **B.** Mutations generated by the exon 5 CRISPR construct. The wild-type sequence is shown with the CRISPR construct sequence highlighted in gray and the PAM site in yellow. The 6F4 allele is a small deletion upstream from the PAM site and the 3F1 allele is a mutation of several basepairs upstream from the PAM site. **C.** The 2F3 allele is a large deletion removing about 2 kb of the *mthl5* gene. PCR analysis of embryonic genomic DNA using primers amplifying the 4.4kb region around *mthl5* is shown. In the 2F3 allele, the amplified fragment is closer to 2kb. **D-D'.** Transcript and protein structure in the *mthl5* alleles. D. The 2F3 allele was sequenced confirming a large deletion in the *mthl5* gene, extending from a site near the beginning of intron 1 and extending through to the CRISPR target site in exon 5. D'. Wild-type *mthl5* encodes a 496 residue protein with a GPCR domain containing the seven transmembrane domains. The proteins encoded by the *mthl5* mutant alleles are shown with the out-of-frame regions in pink. **E.** Topology cartoon of Mthl5 indicated where each mutant allele goes out of frame. Cysteine residues that provide structure to the large extracellular N-terminus of Mthl5 are show in blue. **F.** Variable transcript levels are observed in *mthl5* mutants alleles. As a control, expression of *mthl5* from a non-mutant line from the CRISPR/CAS9 mutagenesis is shown ($mthl5^{1F1}$). Heterozygous embryos are marked with *lacZ* due to a marker insertion on the balancer chromosome TM6BUL.


CGG
CGG
ATTAACCGCAGGACGGTATACGG
GPCR
+49aa
+5aa
+40aa
Extracellular
Cytoplasm
2F3
6F4
3F1
mthl5
mthl5^1F1/TM6BUL
mthl5^3F1/TM6BUL
mthl5^6F4/TM6BUL
mthl5^2F3/TM6BUL
mthl5^1F1
mthl5^3F1
mthl5^6F4
mthl5^2F3

Each mutant line generated by CRISPR was analyzed for *mthl5* transcripts. Heterozygous embryos all showed *mthl5* expression in the salivary gland placode at stage 10 (**Figure 2F**, top panels). Their homozygous siblings (identified by the absence of balancer chromosome *lacZ* expression) retain variable levels of transcript (**Figure 2F**, bottom panels). $mthl5^{3F1}$ and $mthl5^{6F4}$ had approximately wild-type levels of expression, consistent with the small number of nucleotides deleted in each of these alleles. $mthl5^{2F3}$ had considerably lower levels of *mthl5* expression, consistent with the large amount of *mthl5* DNA deleted in this allele.

Loss of *mthl5* causes salivary gland defects

Salivary glands were examined for defects in the *mthl5* CRISPR mutants. All homozygous lines showed irregularities in the apical membrane, as revealed by Crb staining at stage 13 (**Figure 3A**). The main defects were bulges and irregular apical membrane, which is normally smooth and of even diameter in wild-type glands. The penetrance of the mutant phenotypes was reduced relative to the deficiencies, suggesting that something else removed by both deficiencies may also affect salivary gland morphology. To determine the nature of the alleles, the phenotypes of each *mthl5* allele was assessed either in trans to a *mthl5* deficiency or a wild-type chromosome. The quantification of the apical membrane irregularities (**Figure 3A'**) revealed that $mthl5^{2F3}$ behaves as a null allele. The $mthl5^{3F1}$ and $mthl5^{6F4}$ alleles also act as nulls although they may have some mild dominant negative effects, since the homozygotes have more severe defects than when each allele is in trans to a deficiency.

To look more closely at earlier stages when *mthl5* is first expressed and when the salivary glands undergo invagination, embryos were stained with E-Cadherin (E-Cad). E-Cad localizes to the adherens junctions of epithelial cells and is an indicator of apical domain size. In wild-type salivary glands, just after the start of invagination, the cells closest to the invagination pit have

decreased apical size and sharp intense E-Cad outlining the apical membrane (**Figure 3B**). In the *mthl5*^{2F3} mutant salivary glands, the levels of E-Cad appear less intense than in wild-type (**Figure 3B'**). To quantify the E-Cad staining, line scans starting from the invagination pit and moving towards the anterior domain of the placode were performed. Wild-type salivary gland cells had a higher E-Cad pixel intensity than in the *mthl5*^{2F3} cells (**Figure 3C**, compare y-axis of blue and red lines). The peaks of intensity in the *mthl5*^{2F3} cells were also farther apart within three cell lengths of the pit when compared to the same region of wild-type salivary glands. Altogether, these data suggest that, with loss of *mthl5*, the apical domains of cells do not constrict as much as in wild-type, resulting in the less intense Ecad staining at AJs.

Localization and overexpression of Mthl5 in salivary gland cells

To learn where in the cell Mthl5 localizes, antiserum was generated to the long extracellular N-terminal domain of Mthl5. Unfortunately, this antiserum failed to recognize endogenous levels of Mthl5, so it was not useful for examining protein levels in wild-type embryos or in the *mthl5* CRISPR mutants. However, the antiserum did detect over-expressed Mthl5 (**Figure 4A, A'**). Using the *fkh*-Gal4 salivary gland driver, UAS-*mthl5* was expressed in the salivary gland. Prior to invagination, the Mthl5 antiserum recognized Mthl5 in puncta throughout the salivary gland placode **(Figure 4A'**). During invagination (**Figure 4B-B''**), Mthl5 was in still localized in puncta but with strong enrichment at the apical plasma membranes, where it colocalized with the apical Crb protein (yellow, **Figure 4B''**). This apical localization of Mthl5 is consistent with a role for this protein in salivary gland invagination.

Interestingly, at later developmental stages, localization of overexpressed Mthl5 was different. During stage 13, (when endogenous *mthl5* transcripts are no longer detected), Mthl5 still colocalized with apical markers (**Figure 4C''**), but, by stage 14, Mthl5 localized exclusively to

Figure 3: Salivary gland defects are observed with loss of *mthl5*. A-A'. *mthl5* mutant alleles show irregularities in the salivary gland apical membrane at stage 13. The number of irregularities are quantified in A'. **B-B'.** Changes in ECad enrichment are observed in *mthl5* alleles. C-C'. ECad staining in wild-type and *mthl52F3* embryos. The yellow line extends from the invagination pit (left) towards the anterior end of the embryo (arrowhead). Note: the lines are representative and are not the exact portion those quantified in (C). **C.** Quantification of the linescans of WT and *mthl52F3* embryos

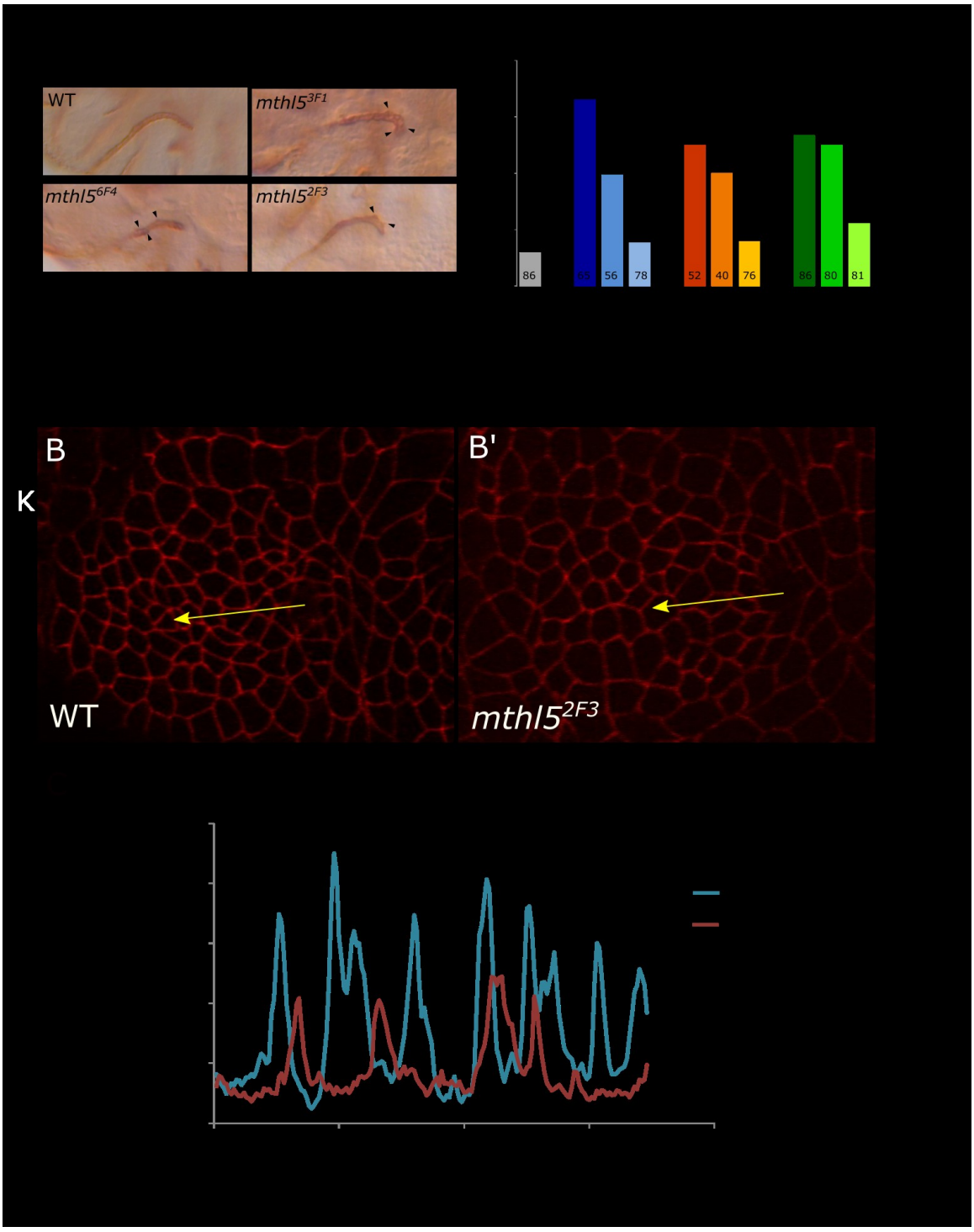
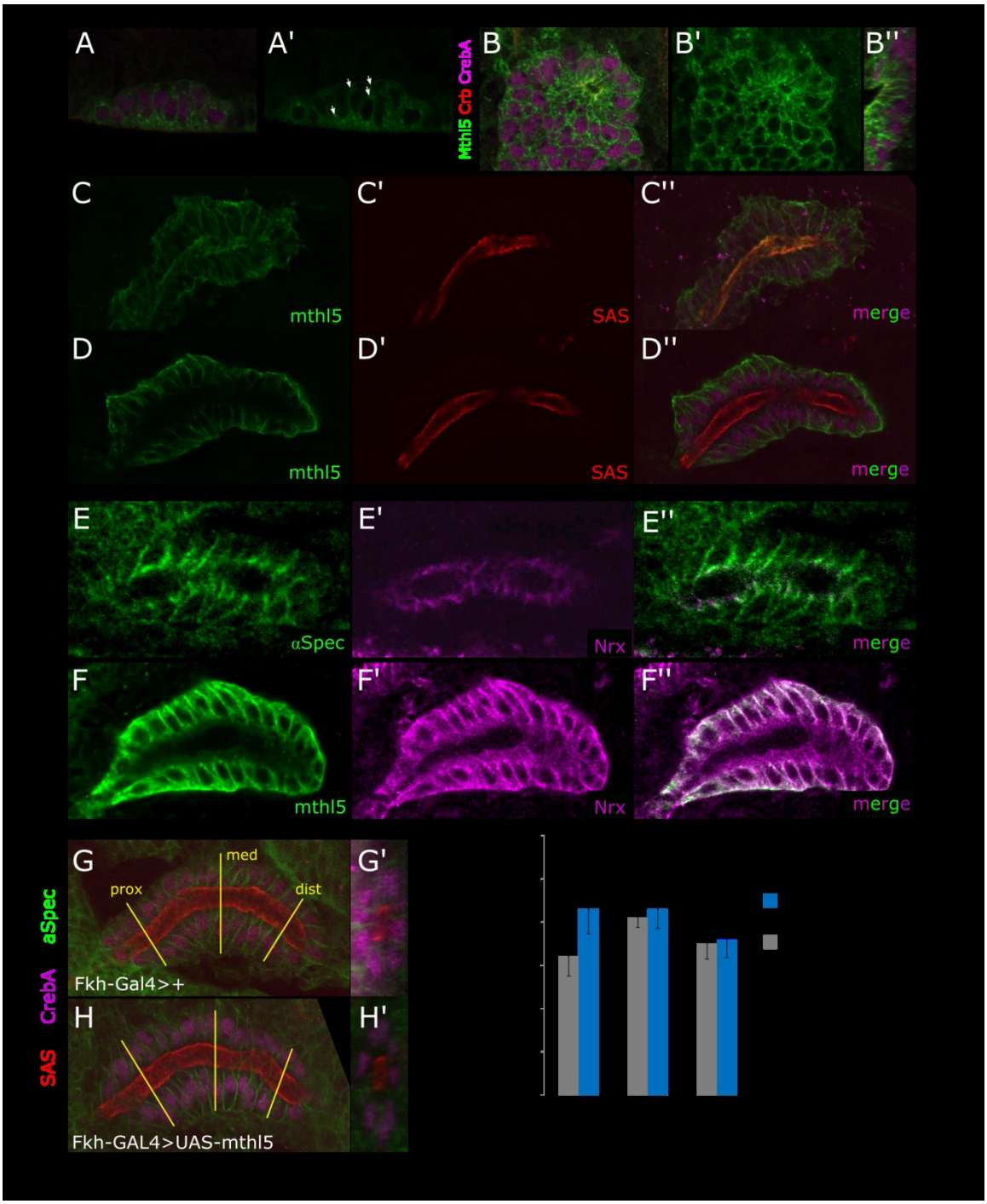

WT
mthl5^3F1
mthl5^6F4
mthl5^2F3
86
65
56
78
52
40
76
86
80
81
K
B
B'
WT
mthl5^2F3

Figure 4. Over-expression of *mthl5* in the salivary gland causes mislocalization of proteins in the basolateral domain. A-B''. Confocal images are shown of *fkh-Gal4* driving *UAS-mthl5* expression in the salivary gland. A-A'. Before invagination, Mthl5 is localized in cytoplasmic puncta and at the membranes (lateral views). B-B'' During invagination, Mthl5 is enriched apically at the invagination pit (ventral views). B'' is an XY orthogonal view of the invaginating gland. **C-D''** Mthl5 translocates from the apical and lateral membranes to the basolateral membranes. C-C'' Stage 13 embryos show Mthl5 in the lateral and apical membranes (C). The apical localization overlaps with SAS staining (C',C'') By stage 14, the Mthl5 staining no longer colocalizes with SAS and is exclusively located at the basolateral membranes. **E-F''.** Over-expression of Mthl5 causes mislocalizatoin of a septate junction protein. E-E'' Wild type Nrx is located apically to the lateral αSpec staining. F-F''*mthl5* expression driven by *fkh-Gal4* results in Nrx mislocalization (F'). Apical surface protein localization is not affected by *mthl5* overexpression (H). **G-I.** Over-expression of *mthl5* by *fkh-Gal4* increases the number of nuclei around the lumen of the gland but the changes are not statistically significant. Quantification is shown in I.


A
A'
B
B'
B''
Mthl5 Crb CrebA
C
C'
C''
mthl5
SAS
merge
D
D'
D''
mthl5
SAS
merge
E
E'
E''
αSpec
Nrx
merge
F
F'
F''
mthl5
Nrx
merge
G
G'
med
prox
dist
aSpec
CrebA
Fkh-Gal4>+
H
H'
SAS
Fkh-GAL4>UAS-mthl5

basal and lateral membranes (**Figure 4D**) and no longer colocalized with the apical marker SAS (**Figure 4D''**). This apical to basal movement of Mthl5 did not affect localization of SAS (**Figure 4G-H'**); it did, however, cause mis-localization of the septate junction protein Nrx-IV (**Figure 4E-F'**). Invertebrate septate junctions are roughly equivalent to vertebrate tight junctions in that they contain homologous proteins and provide epithelial barrier function. In contrast to tight junctions, which are found just apical to adherens junctions, septate junctions are located just basal to adherens junctions. Late *mthl5* over-expressing salivary glands have Nrx-IV spread throughout the lateral membrane instead of localizing just basal to the adherens junctions as they are in wild-type (**Figure 4E', F'**). The localization of the lateral protein αSpectrin (αSpec) remained unchanged (**Figure 4G, H**). Altogether, these findings suggest that when *mthl5* is expressed during later stages than normal, it localizes to a different domain where it can affect some aspects of cell polarity, perhaps through the recruitment of cytoskeletal proteins whose localization is known to be affected by GPCR signaling (Kunwar et al 2008; Schwabe et al 2005).

The changes in the basolateral domain protein localization observed in the *mthl5* over-expressing salivary glands prompted us to ask if there were any defects associated with *mthl5* overexpression. As the gland invaginates from a placode, the cells rearrange from a plate-like structure of ~144 cells on the embryo surface to an elongated tube, roughly eight cells in circumference and 18 cells long. To determine if SGs overexpressing *mthl5* formed tubes of normal dimensions, the number of nuclei around the tube at a proximal, medial, and distal positions was counted (**Figure 4G-I**). Although there were slightly more cells around the tube in *mtlh5* over-expressing glands (**Figure 4I**), the changes were not statistically significant.

Fog as a ligand for both Mist AND Mthl5

As with most GPCRs, *mthl5* is classified as an orphan receptor since its ligand is unknown. Because Mthl5 is a member of the Mth superfamily, we wondered if any information regarding other Methuselahs could direct us towards a possible ligand. Candidate ligands for only two members of the Mth family have been identified: Sex peptide as a potential ligand for Mth and Folded gastrulation (Fog) as a known ligand for Mist (Mthl1). Mthl5, Mthl14, Mthl15 and Mist do not fall into the large superclade of Mth proteins (**Figure 1A**), so perhaps some of these receptors share Fog as an activating ligand. To determine if Fog could be a ligand for Mthl5, we first looked at its expression pattern. *fog* is expressed in the ventral furrow at Stage 5, as is *mist*, consistent with both genes' role in gastrulation (**Figure 5A**). Interestingly, *fog* is also expressed in the salivary gland during stage 11, coinciding nicely with *mthl5*'s expression pattern. *mist* is not expressed in the salivary gland, suggesting that there is another receptor for Fog in the salivary gland. The best candidate for the salivary gland Fog GPCR is Mthl5 since it is expressed in the salivary gland at the right time. *mthl14*, a Mth family member that separates into the same clade as *mthl5* is not expressed in either the ventral furrow or the salivary gland (*mthl14*) and *mthl15* is not expressed in embryos based on RNAseq data (Flybase). These data were encouraging, and further suggested that Fog could be a potential Mthl5 ligand.

In the ventral furrow, both *fog* and *mist* are controlled by the mesodermal transcription factors Twist and Snail (Twi and Sna; Manning et al 2013; Seher et al 2007). When Mthl5 binds Fog, the $G\alpha_{12/13}$ protein Concerntina (Cta) is activated (Parks and Wieschaus, 1991), which then activates Rho1 via RhoGEF2 activation (**Figure 5B**; Barrett et al 1991; Nikolaidou and Barrett, 2004). Activated Rho1 leads to the cell shape changes that drive gastrulation. Rho1 and RhoGEF2 have previously been shown to be important for salivary gland invagination (Xu et al 2008), although it was unknown how these proteins become activated. *twi* is not expressed in

the salivary gland at stage 12, but the transcription factor Fkh is. Thus, we asked if *mthl5* and/or *fog* expression in the salivary gland is/are regulated by Fkh. Indeed, our previous microarray analysis revealed that *mthl5* transcript levels decreased by 2.4 fold in *fkh* mutant embryos (Maruyama et al 2011). To determine if *fkh* is necessary for *mthl5* expression in the salivary gland specifically, transcript levels of *mthl5* were compared between *fkh* mutant embryos and wild type (**Figure 5G-G'**). *mthl5* transcripts are observed in the salivary gland of *fkh* mutant embryos, although at a lower level than in their heterozygous sibling controls. These data suggest that *fkh* contributes to the salivary gland expression of *mthl5*, but is not absolutely necessary. Similar in situ analysis of *fog* expression in *fkh* mutants has shown that *fog* expression disappears in a *fkh* mutant (SeYeon Chung, unpub.), suggesting that *fkh* is necessary for wild-type levels of both *mthl5* and *fog* expression.

To learn if Fkh could activate expression of *mthl5* and *fog* in ectopic domains, *UAS-fkh* was driven with an *en-Gal4* driver, which is expressed in ectodermal stripes. During stage 12, endogenous *mthl5* is expressed in the salivary gland placode and in a subset of mesodermal cells (**Figure 5C**). In embryos ectopically expressing *fkh* in ectodermal stripes, *mthl5* is also observed in the striped *en* domain (**Figure 5D**). The same striped expression is observed at stage 14 **(Figure 5C', D'**). Similarly, *fog* expression is also observed in stripes in the *en-Gal4>UAS-fkh* embryos (**Figure 5E-F'**). Therefore, Fkh is sufficient to drive expression of both *mthl5* and *fog*.

fog mutants were examined for a possible role in salivary gland development. Loss of *fog* causes major defects in embryonic development since gastrulation fails. In *fog* mutants, salivary glands form, but typically stay at or near the surface of the embryo (**Figure 6B'**), often completely failing to invaginate. Because the failure of salivary glands to invaginate could be linked (at least partially) to the failure of the mesoderm to invaginate, we looked more specifically for a role for salivary gland-expressed *fog.* To do this, salivary gland specific RNAi of

Figure 5: Fkh regulates *mthl5* and *fog* expression in the salivary gland. A. Expression of *fog*, *mist*, *mthl5*, and *mthl14* in stage 5 and stage 11 wild-type embryos. Ventral furrow expression is indicated at stage 5. Closed arrowheads indicate expression, open arrowheads indicate lack of expression. During stage 11, salivary gland expression is indicated in the same way. **B**. Model of the apical constriction pathway in the mesoderm and salivary gland. In the mesoderm, *fog* and *mthl5* expression are downstream of Twi. In the mesoderm, Fog binds Mist (open arrow) and then activates RhoGEF2 through the Gα Cta black arrow). RhoGEF2 then activates Rho, which activates Rho kinase (Rok), which phosphorylates and activates myosin leading to apical constriction. In the salivary gland, we imagine a similar pathway wherein Fog binds Mthl5 (open arrow) and Mthl5 then activates downstream signaling (gray arrows). Our hypothesis is that Fkh controls expression of Fog and Mthl5 in the salivary gland. **C-F'.** Ectopic expression of *fkh* via *en*-Gal4 drives ectopic *mthl5* and *fog* expression. Wild-type *mthl5* is expressed in the salivary gland placode and mesoderm at stage 11 (C). At Stage 14, expression is very limited (C'). When *fkh* is driven in the striped Engrailed (En) domain, expression of *mthl5* is detected in stripes (red arrowheads; D, D'). Much like *mthl5, fog* is expressed in the salivary gland placode in wild-type embryos at stage 12 (E) and at low levels at stage 14 (E'). Ectopic *fog* is detected in stripes when *fkh* is driven by engrailed (F, F'). **G-G'.** *mthl5* expression is detected in *fkh* mutant embryos.

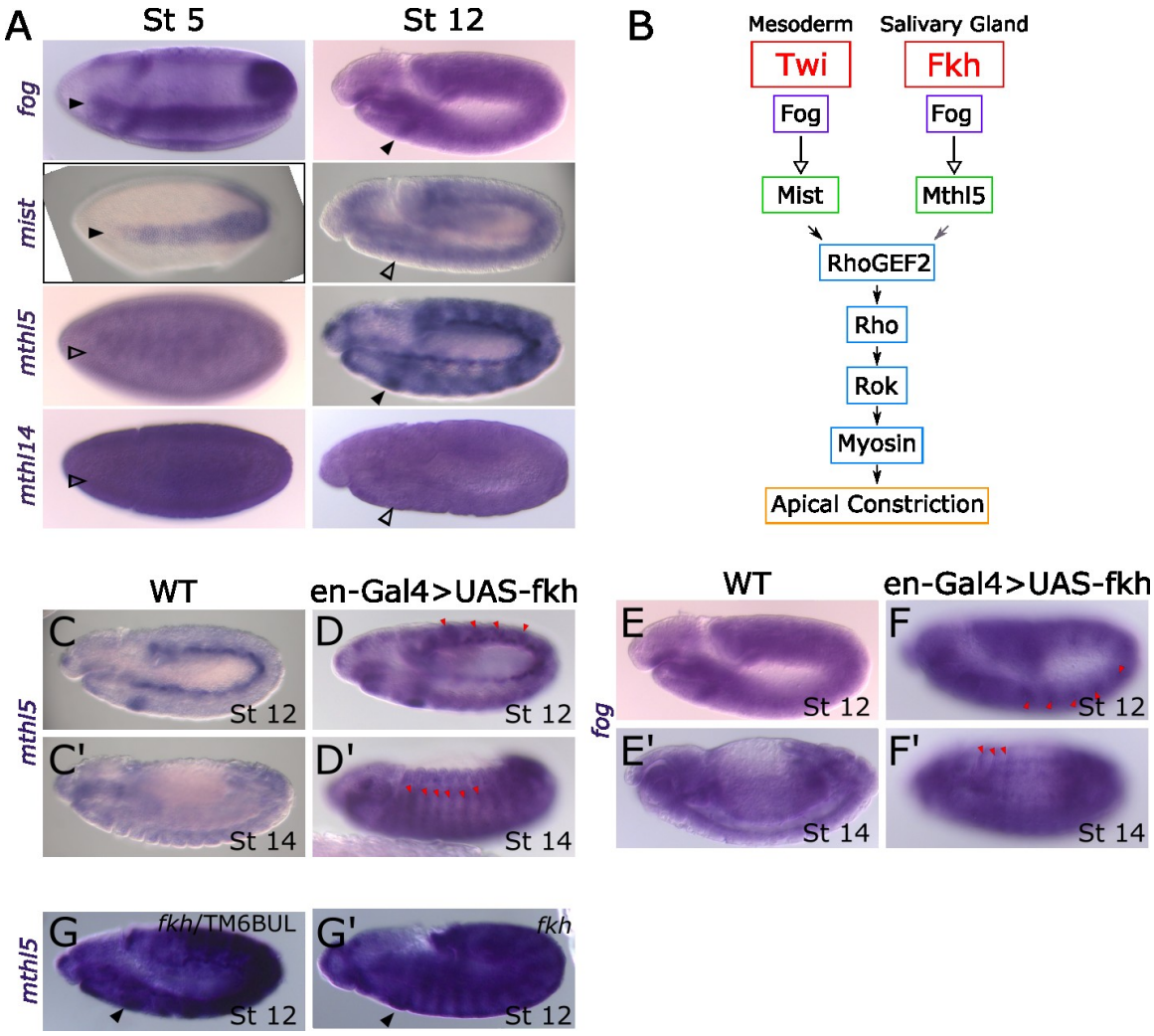

A
St 5
St 12
fog
mist
mthl5
mthl14
B
Mesoderm
Salivary Gland
Twi
Fkh
Fog
Fog
Mist
Mthl5
RhoGEF2
Rho
Rok
Myosin
Apical Constriction
WT
en-Gal4>UAS-fkh
C
St 12
D
St 12
C'
St 14
D'
St 14
mthl5
WT
en-Gal4>UAS-fkh
E
St 12
F
St 12
E'
St 14
F'
St 14
fog
G
fkh/TM6BUL
St 12
G'
fkh
St 12
mthl5

Figure 6. Modulation of *mthl5* and *fog* levels cause salivary gland defects. A-D Knockdown, loss-of-function (LOF), and over-expression of *mthl5* or *fog*. **A-A'** Control embryos showing smooth apical membranes as revealed by Crb staining. **B-C** RNAi of *fog* (B) and *mthl5* (C) both show widened regions of the apical membrane (arrowheads). **B'-C'** Loss-of-function (LOF) alleles of *fog* (B') and *mthl5* (C'). The defects seen in the fog^{s4} allele are quite severe possibly due to accompanying defects in gastrulation. The *mthl5* null allele shows rough and widened apical membranes. **B''-C''** Over-expression of *fog* (B'') and *mthl5* (C'') causes widened apical membranes. **D.** Over-expression of both *fog* and *mthl5* together also causes widened apical membranes. **E.** Quantification of the defects seen in A-D. **F.** Mix-expression of *mthl5* by Twi-Gal4 does not affect salivary gland migration.

A
WT
B
fog-RNAi
B'
fog^S4
B''
UAS-fog
A'
Fkh-Gal4>WT
C
mthl5-RNAi
C'
mthl5^2F3
C''
UAS-mthl5
D
UAS-fog, UAS-mthl5
F
Twi-Gal4>UAS-mthl5
94
109
147
96
12
86
29
60
74

fog was carried out. Knockdown of *fog* in the salivary gland using the *fkh*-Gal4 driver causes bulges and irregularities in the salivary gland much like those observed with *mthl5* CRISPR mutations and *mthl5* deficiencies (**Figure 6B**). Very similar phenotypes were observed with *mthl5* knockdown using *fkh-Gal4* to drive *mthl5* RNAi (**Figure 6C**). These findings are fully consistent with the hypothesis that Fog is the Mthl5 ligand in the salivary gland.

When *mthl5* is over-expressed in the salivary gland, mild apical membrane irregularities are observed (**Figure 6C''**), whereas over-expression of *fog* causes the lumen to be wider than normal (**Figure 6B''**). Over-expression of both *mthl5* and *fog* in the salivary gland causes defects similar to those seen in *fog* over-expression (**Figure 6D**), suggesting that Fog is the limiting factor. Quantification of these data is shown in Figure 5E. The wider lumens observed with Fog overexpression in the SG could suggest that more cells internalize at earlier stages, consistent with the idea that Fog acts through Mthl5 to drive apical constriction and salivary gland invagination.

Finally, we asked if there were any defects associated with ectopic expression of *mthl5* in the ventral furrow using *twi-Gal4*. The goal of this experiment was to ask if providing ectopic *mthl5* in the *mist* domain would block Mist signaling by sopping up Fog. *Twi-Gal4>UAS-mthl5* embryos develop normally and salivary gland invagination is not affected (**Figure 6L**). Thus, either endogenous Mist is still able to bind sufficient Fog to drive gastrulation in the background of *mthl5* mis-expression, or Mthl5 may activate the same signaling pathway as Mist to drive mesodermal invagination. Driving *mthl5* in the ventral furrow in a *mist* mutant background should resolve this potential ambiguity.

Discussion

mthl5 is a member of the Methuselah family of Secretin GPCRs. The Methuselah family is the third largest family of GPCRs encoded in the *Drosophila* genome, and yet only two of its sixteen members have been characterized to any extent. *mist*, or *mthl1*, is the best characterized Mth family member, and has been identified as the receptor for Fog. *mist* and *fog* are expressed in an overlapping region in the ventral furrow, and loss of either *mist* or *fog* causes defects in gastrulation. Because defective gastrulation is an easily recognized phenotype, many players in the Fog pathway had been identified before *mist* was characterized. The first clue that a GPCR was vital to gastrulation was the discovery that Concertina, a Gα subunit, also shows gastrulation defects (Parks and Wieschaus, 1991). The GPCR kinase Gprk2 also affects gastrulation (Fuse et al 2013), as does the Gα GEF Ric-8 (Peters and Rogers, 2013). Downstream of Cta, it was shown that RhoGEF2 strongly affected ventral furrow formation (Barrett et al 1997), and that apical myosin levels were lower in a *RhoGEF2* mutant (Nikolaidou and Barrett, 2004). Further work showed that Rho kinase was activated downstream of Fog signaling, and that Fog caused myosin to localize apically prior to the observed cell shape changes (Dawes-Hoang et al, 2005). Mist provided the link between Fog and the Cta-RhoGEF-Rho-Rok-Myosin signaling pathway.

Here, we show that many of the same signaling pathway components are expressed in the salivary gland. *fog* was known to cause defects in embryonic salivary glands (Lammel and Saumweber, 2000), although the defects are not specific to the salivary gland itself. Rho and Rok were known to be involved in gland invagination, as is a negative regulator of RhoGAPS (Xu et al, 2008; Kolesnikov and Beckendorf, 2007). Myosin is also dynamically involved in gland invagination (Roper, 2012). Each placode is surrounded by a myosin cable, and myosin changes from a cortical to a medial position in cells near the invagination pit. Moreover, cells in the

salivary gland placode undergo cell shape changes similar to those observed in ventral furrow formation (Myat and Andrew, 2000b). In *fkh* mutant embryos, these cell shape changes do not occur, reminiscent of the absence of cell shape changes in the ventral furrow in *snail* and *twist* mutant embryos (Grau et al 1984). The link in the salivary gland between *fkh* and the cell shape changes and between Fog, Rho, and myosin was unknown until this work.

Of the sixteen Mth family members, four separate away from the main Mth superclade. Mist and Mthl5 are both in the Mth family offshoot. *mthl5* is not expressed in the ventral furrow, as *fog* and *mist* are, but it is expressed in the salivary gland during gland invagination and dorsal movement. Interestingly, spatiotemporal expression of *fog* matches that of *mthl5* in the salivary gland, and the known Fog ligand, encoded by *mist* is not expressed in this region. Our studies indicate that Fog is also a ligand of Mthl5. We created both loss-of-function alleles via CRISPR/Cas9 as well as constructs for *mthl5* over-expression. Loss of *fog* and loss of *mthl5* cause salivary gland defects and RNAi of either give nearly identical defects. *fog* mutants have severe defects due to incomplete gastrulation – the embryos appear twisted and the salivary glands form but do not invaginate. While the failure of gland invagination could be due to the gross morphological defects in the embryo, it could also be from altered signaling from the mesoderm to the salivary glands. It would be interesting to determine if loss of *mist* and *mthl5* phenocopies a loss of *fog* for both mesoderm and salivary gland formation. Over-expression of *fog* also causes enlarged salivary gland lumens, where over-expression of *mthl5* causes only mild defects, suggesting that ligand activation is necessary. Both *mthl5* and *fog* salivary gland expression go down in *fkh* and can be activated by ectopic *fkh* expression, further implicating Mthl5 in the Fkh/Fog signaling pathway required to internalize the salivary gland.

Close examination of loss of *mthl5* showed that E-Cadherin dynamics are altered. During salivary gland migration, E-Cadherin is strongly localized to the membrane, where it can be used

to track the cell shape changes that occur prior to salivary gland invagination (Booth et al 2014). In *mthl5* mutants, E-Cadherin is not as strongly localized to the membrane, and the apical domains are larger than in wild-type salivary gland placodes. In dorsal fold formation, invagination of the cells is preceded by a ventral shift of E-Cad (Wang et al 2013), and a similar observation has been made in the salivary gland, at least for the first cell that internalizes (SangJoon Kim, unpublished results). Lack of apical constriction is also observed when the microtubule network is disrupted in the salivary gland (Booth et al 2014), a phenotype mediated by altered myosin dynamics in the gland placode. Apical constriction is also affected by loss of *fkh* (Seyeon Chung, unpublished results). Altogether, these data support a hypothesis that *mthl5* plays a role in the cell shape changes that drive salivary gland internalization. Further investigations into E-cad localization/accumulation and myosin dynamics in the *mthl5* mutant are being done now to link *mthl5* to these processes.

A final test of whether Mthl5 is the Fog receptor in the salivary gland will be done by expressing *mthl5* in S2 cells and then exposing these cells to secreted Fog protein. This approach is similar to that used by Manning et al (2013) to identify Mist as a Fog receptor. However, so far, we have been unable to obtain reliable *mthl5* expression in S2 cells. We will also over-express *fog* in a $mthl5^{2F3}$ embryo to see if the *fog* over-expression phenotypes are mediated through Mthl5. We are now attempting to express *mthl5* and *fog*, alone and together in the wing imaginal disc, a system often used to examine apical constriction, cell shape changes, and folding (Zimmerman et al 2010). Our hypothesis is that expression of both genes will lead to cell shape changes, whereas expression of each component alone will not. We hope that these experiments will conclusively demonstrate that Mthl5 is a receptor for Fog, and therefore fully implicate *mthl5* in the signaling pathway driving salivary gland invagination.

It should be noted that even with loss of *mthl5* and knock-down of *fog*, the salivary gland still invaginates. This finding suggests that redundant mechanisms exist within the salivary gland to drive invagination. Another GPCR, *Tre1*, is expressed in the salivary gland at stage 12, and may function with *mthl5* during this stage of development. Moreover, apical constriction is not the only mechanism necessary for internalization. Over-expression of Crb in the salivary gland prevents apical constriction, but the glands invaginate, albeit with defects in their morphology (SeYeon Chung and Sangjoon Kim, unpublished). In the trachea, apical constriction is paired with an additional cell division that drives internalization (Kondo and Hayashi, 2013). *fkh* is the only known component to completely block salivary gland migration, so the characterization of other Fkh targets as well as studies of the interaction of these targets with *mthl5* and *fog* would add insight to how multiple mechanisms drive salivary gland internalization.

REFERENCES

Andrew D.J., Baig A., Bhanot P., Smolik S.M., and Henderson K.D. (1997). The *Drosophila* dCREB-A gene is required for dorsal/ventral patterning of the larval cuticle. *Development*, **124**, 181–193

Baena-Lopez, L.A., Alexandre, C., Mitchell, A., Pasakarnis, L., and Vincent, J-P. (2013) Accelerated homolgous recombination and subsequent genome modification in *Drosophila*. *Development*, **140**, 4818-25.

Barrett, K., Leptin, M., and Settleman, J. (1997). The Rho GTPase and a putative RhoGEF mediate a signaling pathway for the cell shape changes in Drosophila gastrulation. *Cell*, **91**(7), 905–915.

Booth, A. J. R., Blanchard, G. B., Adams, R. J., and Röper, K. (2014). A Dynamic microtubule cytoskeleton directs medial actomyosin function during tube formation. *Developmental Cell*, **29**(5), 562–576.

Costa, M., Wilson, E. T., and Wieschaus, E. (1994). A Putative Cell Signal Encoded by the folded gastrulation Gene Coordinates Cell Shape Changes during Drosophila Gastrulation, *Cell*, **76**(6), 1075-1089.

Cotton, M., and Claing, A. (2009). G protein-coupled receptors stimulation and the control of cell migration. *Cellular Signaling*, **21**(7), 1045–1053.

Cvejic, S., Zhu, Z., Felice, S. J., Berman, Y., and Huang, X.-Y. (2004). The endogenous ligand Stunted of the GPCR Methuselah extends lifespan in Drosophila. *Nature Cell Biology*, **6**(6), 540–6.

Dawes-Hoang, R. E., Parmar, K. M., Christiansen, A. E., Phelps, C. B., Brand, A. H., and Wieschaus, E. F. (2005). Folded Gastrulation, Cell Shape Change and the Control of Myosin Localization. *Development*, **132**(18), 4165–4178.

Fox, R. M., Vaishnavi, A., Maruyama, R., and Andrew, D. J. (2013). Organ-specific gene expression: the bHLH protein Sage provides tissue specificity to Drosophila FoxA. *Development*, **140**(10), 2160–71.

Fuse, N., Yu, F., and Hirose, S. (2013). Gprk2 adjusts Fog signaling to organize cell movements in Drosophila gastrulation. *Development*, **140**(20), 4246–55.

Grau, Y., Carteret, C., and Simpson, P. (1984). Mutations and chromosomal rearrangements affecting the expression of *snail*, a gene involved in embryonic patterning in Drosophila melanogaster. *Genetics,* **108**(2), 347-360.

Greig, S., and Akam, M. (1993). Homeotic genes autonomously specify one aspect of pattern in the Drosophila mesoderm. *Nature ,* **362**(6421), 630-632.

Hacker, U., and Perrimon N. (1998). DRhoGEF2 encodes a member of the Dbl family of oncogenes and controls cell shape changes during gastrulation in Drosophila. *Genes Dev.* **12**, 274–284.

Harmar, A J. (2001). Family-B G-protein-coupled receptors. *Genome Biology*, **2**(12).

Henderson, K.D., and Andrew, D.J. (2000). Regulation and function of Scr, exd, and hth in the Drosophila salivary gland. *Dev. Biol.* **217**(2), 362-374.

Ja, W. W., Carvalho, G. B., Madrigal, M., Roberts, R. W., and Benzer, S. (2009). The Drosophila G protein-coupled receptor, Methuselah, exhibits a promiscuous response to peptides. *Protein Science : A Publication of the Protein Society*, **18**(11), 2203–8.

Ja, W. W., West, A. P., Delker, S. L., Bjorkman, P. J., Benzer, S., and Roberts, R. W. (2007). Extension of Drosophila melanogaster life span with a GPCR peptide inhibitor, **3**(7), 415–419.

Kobilka, B. K. (2007). G protein coupled receptor structure and activation. *Biochimica et Biophysica Acta*, **1768**(4), 794–807.

Kolesnikov, T. and Beckendorf, S, K. (2007). 18 wheeler regulates apical constriction of salivary gland cells via the Rho-GTPase signaling pathway. *Developmental Biology*, **307**(1), 53-61.

Kondo, T., and Hayashi, S. (2013) Mitotic cell rounding accelerates epithelial invagination. *Nature* **494**(7435),

Kunwar, PS, Sano, H, Renault, AD, Barbosa, V, Fuse, Naoyiku, and Lehmann, R. (2008). Tre1 GPCR initiates germ cell transepithelial migration by regulating *Drosophila melanogaster* E-Cadherin. *JCB* **183**(1), 157-168.

Lammel, U., and Saumweber, H. (2000). X-linked loci of Drosophila melanogaster causing defects in the morphology of the embryonic salivary glands. *Dev. Genes Evol.,* **210**(11), 525--535.

Lehmann, R., and Tautz, D. (1994). In situ hybridization to RNA. *Mehtods Cell Biol* **44**, 575-98.

Lin, Y. J., Seroude, L., and Benzer, S. (1998). Extended life-span and stress resistance in the Drosophila mutant methuselah. *Science*, **282**(5390), 943–946.

Manning, A. J., Peters, K. a, Peifer, M., and Rogers, S. L. (2013). Regulation of epithelial morphogenesis by the G protein-coupled receptor mist and its ligand fog. *Science Signaling*, **6**(301).

Maruyama, R., Grevengoed, E., Stempniewicz, P., and Andrew, D. J. (2011). Genome-wide analysis reveals a major role in cell fate maintenance and an unexpected role in endoreduplication for the Drosophila FoxA gene fork head. *PLoS ONE*, **6**(6).

Myat, M. M., and Andrew, D. J. (2000a). Fork head prevents apoptosis and promotes cell shape change during formation of the Drosophila salivary glands. *Development*, **127**(19), 4217–4226.

Myat, M. M., and Andrew, D. J. (2000b). Organ shape in the Drosophila salivary gland is controlled by regulated, sequential internalization of the primordia. *Development*, **127**(4), 679–691.

O'Hayre, M., Vazquez-Prado, J., Kufareva, I., Stawiski, E.W., Handel, T.M., Seshagiri, S., and Gutkind, J.S. (2013). The emerging mutational landscape of G-proteins and G-protein coupled receptors in cancer. *Nat Rev Cancer*, **13**(6), 412-424.

Parks, S., and Wieschaus, E. (1991). The Drosophila Gastrulation Encodes a Ga-like Protein Gene concertina encodes a Ga-like Protein. *Cell*, **64**(2), 447–458.

Patel, M. V., Hallal, D. A, Jones, J. W., Bronner, D. N., Zein, R., Caravas, J., Zahabiya, H, Friedrich, M, and Vanberkum, M. F. A. (2012). Dramatic expansion and developmental expression diversification of the methuselah gene family during recent Drosophila evolution. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*, **318**(5), 368–87.

Peters, K.A., and Rogers, S. L. (2013). Drosophila Ric-8 interacts with the Gα12/13 subunit, Concertina, during activation of the Folded gastrulation pathway. *Molecular Biology of the Cell*, **24**(21), 3460–71.

Röper, K. (2012). Anisotropy of Crumbs and aPKC Drives Myosin Cable Assembly during Tube Formation. *Developmental Cell*, **23**(5), 939–953.

Schwabe, T., Bainton, R.J., Fetter, R.D., Heberlein, U., and Gaul, U. (2005). GPCR signaling is required for blood-brain barrier formation in drosophila. *Cell*, **123**, 133–44.

Seher, T. C., Narasimha, M., Vogelsang, E., and Leptin, M. (2007). Analysis and reconstitution of the genetic cascade controlling early mesoderm morphogenesis in the Drosophila embryo. *Mechanisms of Development*, **124**(3), 167–179.

Vining, M. S., Bradley, P. L., Comeaux, C. A., and Andrew, D. J. (2005). Organ positioning in Drosophila requires complex tissue-tissue interactions. *Developmental Biology*, **287**(1), 19–34.

Wang, Y-C., Khan, Z., Kaschube, M., and Wieschaus, E.F. (2012). Differential positioning of adherens junctions is associated with iniation of epithelial folding. *Nature*, **484**, 390-393.

Weiss, J.B., K.L. Suyama, H.H. Lee, and M.P. Scott. (2001). Jelly belly: a Drosophila LDL receptor repeat-containing signal required for mesoderm migration and differentiation. *Cell*, **107**, 387-98.

Xu, N., Keung, B., and Myat, M. M. (2008). Rho GTPase controls invagination and cohesive migration of the Drosophila salivary gland through Crumbs and Rho-kinase. *Developmental Biology*, **321**(1), 88–100.

Yona, S., Lin, H. H., Siu, W. O., Gordon, S., and Stacey, M. (2008). Adhesion-GPCRs: emerging roles for novel receptors. *Trends in Biochemical Sciences*, **33**(10), 491–500.

Zimmerman, S. G., Thorpe, L. M., Medrano, V. R., Mallozzi, C. A., and McCartney, B. M. (2010). Apical constriction and invagination downstream of the canonical Wnt signaling pathway require Rho1 and Myosin II. *Developmental Biology*, **340**(1), 54–66.

Chapter 5

Characterization of the Fox Family Transcription Factor FoxL1 in Drosophila Embryogensis

ABSTRACT

Members of the Fox family of transcription factors have been well-characterized for their role in several developmental processes. In *Drosophila*, there are nineteen Fox family members. *foxL1*, previous known as *fd64a*, is expressed in several different types of musculature during embryonic development. In later stages of embryonic development, we discovered that *foxL1* is expressed in Ventral Intersegmental 5 (VIS5), a relatively undescribed muscle found in all three of the thoracic segments. Because the salivary gland closely associates with this muscle during lates stages of posterior migration, we asked if FoxL1 and its targets contribute to salivary gland guidance. To explore this hypothesis, we generated a *foxL1* knockout via homologous recombination, a null allele via CRISPR/Cas9, as well as a *foxL1* over-expression construct. Loss of *foxL1* does not overtly affect salivary gland migration, but over- and mis-expression of *foxL1* cause a variety of salivary gland irregularities. Moreover, mis-expression of *foxL1* in all muscles also causes gross muscle abnormalities. This raised the possibility that *foxL1* affects salivary gland migration by disrupting the scaffolding that the salivary gland uses as a track for migration and by providing an ectopic source of signaling molecules. To more closely examine potential defects in the musculature, we created an mCherry knock-in allele of *foxL1*, which I am currently characterizing. We also identified s*ema2a*, a known secreted signaling molecule, as a target of FoxL1. We plan to perform a microarray using our *foxL1* null alleles to identify more targets that could affect salivary gland migration and muscle morphology.

INTRODUCTION

Cell migration is a critical process that occurs throughout life; it can involve single cells, groups of cells or even entire organs. For example, during embryogenesis, neural crest cells migrate long distances as individuals to ultimately populate many different tissues and organs in response to normal developmental cues. Cells also migrate as a population in order to close wounds. Whereas both of these processes would be considered beneficial to an organism, cell migration can also be detrimental, such as when single cancer cells or clumps of cancer cells metastasize to invade different tissues. To simplify studying this complicated process, the *Drosophila* salivary gland provides a unique model system. Each paired gland comprises approximately 140 epithelial cells, which form tubes from plate or placode of polarized primordial cells. Salivary gland cells neither divide nor die during the entire process of morphogenesis. Moreover, the salivary gland migrates as a fully polarized intact tissue to arrive at its final position in the embryo. These features, as well as the ease of genetic manipulation, make the salivary gland an ideal model for understanding the complex process of cell/tissue migration.

The migration of the salivary gland has been described in detail (Chung et al., 2014). Each salivary gland starts out as a placode of polarized epithelial cells on the surface of the embryo. The cells undergo apical constriction to invaginate into the embryo and form a monolayered epithelial tube, with the apical surfaces facing the lumen and the basal surfaces contacting other tissues. The salivary gland moves dorsally as it internalizes until reaching a turning point during mid-embryogenesis when it reorients and migrates posteriorly to reach its final position by the end of embryogenesis. Integrin signaling is essential for migration of the salivary gland (Bradley et al., 2003) and previous work from our lab has shown that the salivary gland contacts multiple distinct tissues during its posterior migration, including the visceral mesoderm and fat body at

early stages (Vining et al., 2005). Salivary gland migration is affected by several signaling pathways. Netrin, expressed in the central nervous system and visceral mesoderm, acts through its salivary gland expressed receptor Frazzled to attract the gland (Kolesnikov and Beckendorf, 2005). Conversely, Slit, secreted by the midline glia, interacts with Robo, expressed in the salivary gland, to act as a repellent. The PDGF receptor and its ligands are also involved in salivary gland placement (Harris et al 2009). However, much remains to be discovered regarding the ligands, receptors, downstream signaling pathways, and relevance of tissue-tissue contacts in guiding salivary gland movement.

The Fox family transcription factors play major roles in the development in all higher organisms (Weigel and Jackle, 1990; Hannenhalli and Kaestner, 2009). Fox proteins (forkhead box) are members of a well-conserved family of winged-helix transcription factors with many diverse functions, ranging from regulation of organ development and growth, to vocal learning (Lee et al 2005; Haesler et al 2004). These proteins have also been implicated in cancer metastasis where their transcriptional targets include cell cycle regulators, growth factors, and cell adhesion molecules (Myatt and Lam, 2007). There are 19 Fox family members in *Drosophila*, 44 in mouse, and 50 in human (Jackson et al, 2010). Fkh, the single *Drosophila* member of the FoxA protein subfamily, was the first Fox gene cloned (Weigel et al 1989). In the salivary gland, *fkh* is expressed early in the gland placode and continues to be expressed for as long as the organ persists (Myat and Andrew, 2000). Loss of *fkh* results in salivary gland cell death; when the glands are kept alive, the cells in the *fkh* mutants fail to undergo apical constriction and the salivary glands do not invaginate. Whereas *fkh* is unable to specify glands by itself, it does maintain its own expression and expression of many genes that implement the salivary gland cell fate choice (Maruyama et al 2011). *fkh* is also expressed the hindgut, Malpighian tubules, and proventriculus (among other tissues), and induces expression of salivary gland specific

genes by collaborating with the salivary gland specific transcription factor Sage, with some contribution from the Senseless transcription factor (Fox et al 2013). In mammals, FoxA is considered a pioneer transcription factor because it binds chromatin early in an ATP-independent manner and then attracts other transcription factors to those sites (Zaret and Carroll, 2011), a role not supported by studies in flies (Abrams et al., 2006).

The 18 other *Drosophila* Fox genes show dynamic expression throughout development (Lee and Frasch, 2004). The expression pattern of *fd64a* is particularly intriguing due to its expression in the somatic visceral muscles that directly contact the salivary gland during late embryonic stages. The closest mouse homologue of Fd64a is FoxL1 (Figure 1A), which is expressed in the mouse mesenchyme and is necessary for proper gut development (Katz et al 2004). FoxL1 regulates gut morphology by limiting the amount of Syndecan-1 and Perlcan secreted by the mesenchyme, which in turn regulates levels of Wnt signaling in the gut epithelium (Perreault et al 2001). Because the salivary gland uses multiple inputs and signals to direct its journey to its final position in the embryo, we asked if *fd64a* plays a role in salivary gland migration or placement.

MATERIALS AND METHODS

Fly Strains

The *FoxL1* knockout line was created via homologous recombination, as described by Gong and Golic (2003). The *FoxL1* CRISPR sites were chosen using the flyCRISPR Optimal Target Finder (http://flycrispr.molbio.wisc.edu/tools). Oligos were ligated into the pU6-BSA-gRNA plasmid (Baena-Lopez et al, 2013) after BsaI digestion. p-mCh-C1 was a gift from the Inoue lab (Johns Hopkins University School of Medicine). Constructs were injected into lines w[1118]; PBac{y[+mDint2]=vas-Cas9}VK00027 flies by Rainbow Transgenic Flies, Inc (Camarillo, CA). UAS-fd64a was generated by Gateway cloning (Invitrogen, Carlsbad, CA) into the pTW UAS vector. This construct were then injected into *w1118* flies by Rainbow Transgenic Flies, Inc (Camarillo, CA). *Df(3L)BSC369, Df(3L)BSC368*, 5053-Gal4, and caps^{02937} were obtained from the Bloomington Stock Center (Bloomington, Indiana). UAS-Sema2a-TM-GFP was a gift from the Kolodkin lab (Wu et al 2011). *twi*-Gal4 (Greig and Akam, 1993), SNS-Gal4 (Kocherlakota et al.,2008), and RP298-Gal4 (Menon and Chia, 2001) were obtained from Elizabeth Chen's Lab (Johns Hopkins University School of Medicine). *bap*-Gal4 was from Manfred Frasch (University of Erlangen-Nuremberg). *fkh*-Gal4 was created by our laboratory (Myat and Andrew, 2000). Homozygous lethal lines were balanced over Cyo-Ftz-LacZ or TM6-Ubx-LacZ embryos, and mutant chromosomes were identified by lacZ staining.

Fd64a antibody generation

The full-length ORF of *fd64a* was cloned in-frame into the pET15b vector (GE Novagen) using the In-Fusion HD Cloning Kit (Clontech). The construct was transformed

into *E Coli*. Protein was induced during log growth phase with 0.1M IPTG and grown over-night at 16^{o}C. Inclusion body preps were performed to isolate the induced protein and injected into rats (Covance).

Immunohistochemistry and in situ hybridizations

In situ hybridization and immunohistochemistry were performed as previously described (Lehmann and Tautz, 1994; Reuter et al., 1990). An *mcherry in situ* probe was made by cloning the genomic fragment into pCR-TOPO-II vector (Invitrogen). Antibody concentrations used in this study are as follows: mαDm0 (ADL1011:1000; Drosophila Studies Hybridoma Bank, Iowa City, IA), RtαFoxL1 (1:1000), RbαSG2 (1:500), RtαCrebA (1:1,000; Andrew et al., 1997), RbαForkhead (1:2000; a gift from S. Beckendorf, Berkeley, CA, USA), RbαMHC (1:500; Kiehart and Feghali, 1986), mαSema2a (1:250, DSHB), RtαTropomyosin (ab50567, 1:100; Abcam), βPS, RbαVg (1:10; Sean Carroll, University of Wisconsin, Madison), GPαSage (1:500; Fox et al 2013), RbαCrebA (1:10,000; Andrew et al., 1997), RbαSAS (1:500; a gift from D. Cavener, Penn State University, PA, USA), RbαGFP (1:500; Invitrogen, mααSpectrin (1:2; DSHB), mαCrumbs (1:100, DSHB), RtαDE-Cad (1:10, DSHB), and mαβgalactosidase (Promega, 1:10,000). Vg staining was performed overnight at room temperature. The FoxL1 signal was amplified through biotinyl tyramide amplification (TSA Biotin System, Perkin Elmer). All secondary antibodies (Vector Labs and Molecular Probes) were used at a 1:500 dilution. HRP images were developed using the Vectastain ABC kit (Vector Laboratories). Confocal images were obtained using a Zeiss LSM 510 or 700 Meta confocal microscope, using a Plan-Neofluor 63x and 100x, 1.3 oil objective and the Zeiss LSM software. All other

images were obtained using a Zeiss Axiophot microscope configured with a Nikon Coolpix 4500 digital camera or a Janoptik ProgResC14 Plus opitical imaging system. Images were taken using a Plan-Neofluor 20x, 0.50 objective. Images were rotated and cropped using Inkscape, an open-source vector graphic editor. All images were obtained at room temperature.

RESULTS

FoxL1 is expressed in a subset of somatic muscles that contact the migrating salivary gland

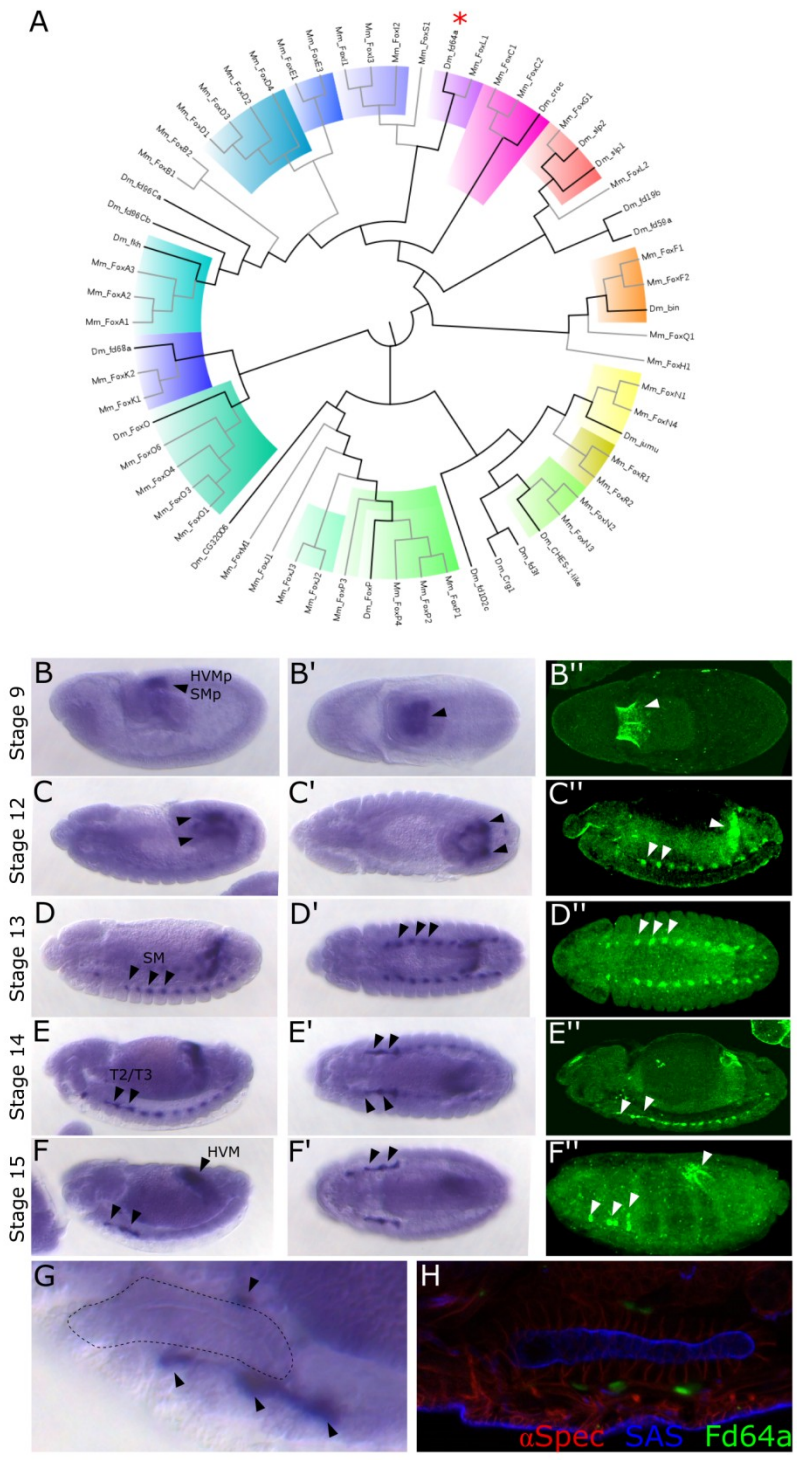
Fd64a is one of 19 Fox family members in *Drosophila* (**Figure 1A**, dark lines). Fd64a is most closely related to mammalian FoxL1 and will heretofore be referred to as *FoxL1*. Like many of the Fox family members, *FoxL1* has a dynamic expression pattern during embryonic development (**Figure 1B-F'**). At stage 9, *FoxL1* is expressed in somatic mesodermal precursors (SMP) and hindgut visceral mesoderm (HVM) precursors (**Figure 1B-B'**). The SMP cells migrate anteriorly through stage 12 (**Figure 1C**), and by stage 13, *FoxL1* is expressed in ten bilateral clusters of somatic mesoderm (**Figure 1D**). The HVM expression of *FoxL1* continues for the duration of embryonic development (**Figure 1C-F'**). *FoxL1* is also expressed in a small cluster of cells in the head region (**Figure 1D-E**). During stage 14, *FoxL1* is upregulated in a group of cells in the second and third thoracic segment; expression in these cells continues through stage 15 and is subsequently lost (**Figure E-F'**). The salivary gland migrates directly between two areas of *FoxL1* expression during stage 15 (**Figure 1G**).

An antibody to full-length FoxL1 was generated and used to stain embryos; the staining pattern with the antiserum recapitulates the *in situ* expression pattern. Consistent with the role of FoxL1 as a transcription factor, Fox1 staining is nuclear. As seen with the RNA, the salivary gland migrates in direct contact with FoxL1 expressing cells (**Figure 1H**). This expression pattern suggested that *FoxL1* may influence salivary gland placement.

Loss of FoxL1 does not overtly affect salivary gland placement

Deficiency lines removing *FoxL1* were obtained to ask if *FoxL1* has a role in salivary gland placement. Wild-type salivary glands at stage 15 have a smooth and even apical membrane, which can be seen with a variety of apical membrane and junctional markers that localize either

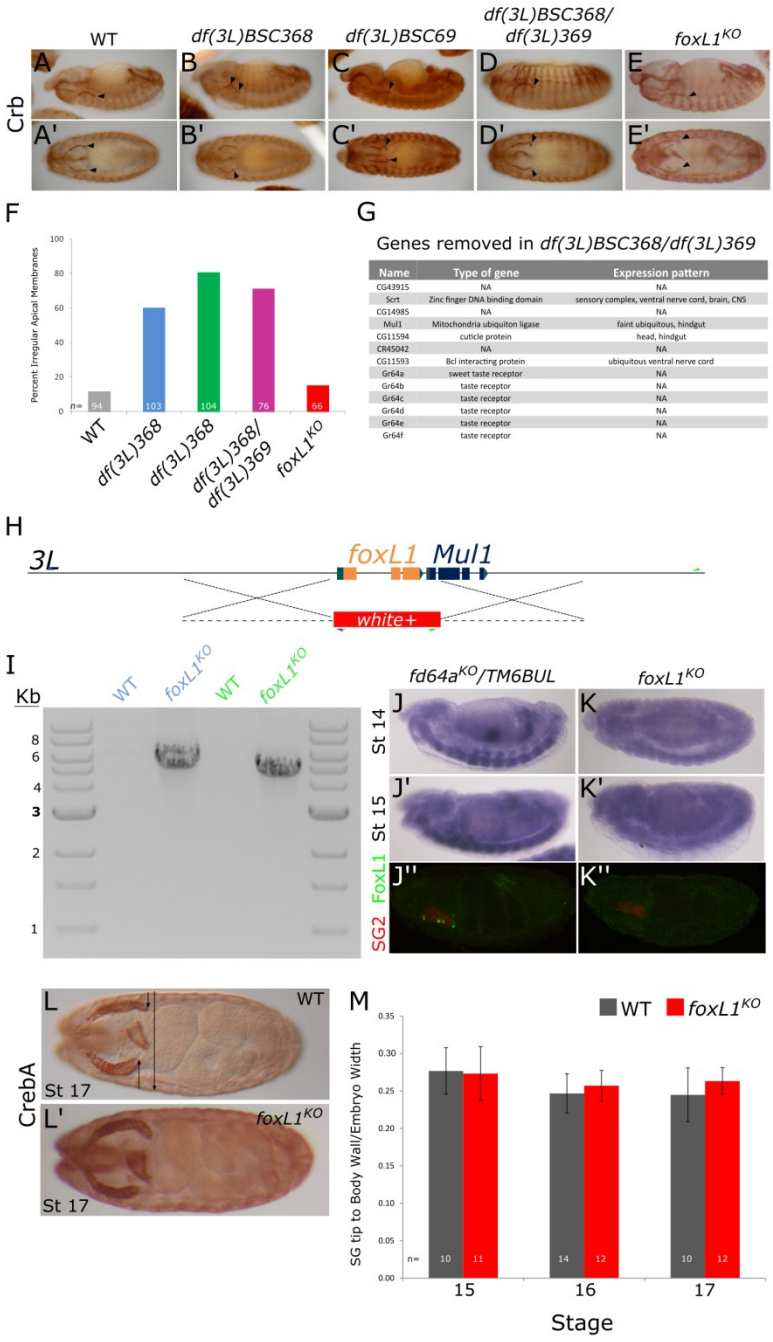
Figure 1: *FoxL1* is a member of the Fox family of transcription factors and it expressed throughout the *Drosophila* embryo. **A**. Phylogenetic tree of mouse (Mm) and fly (Dm) Fox transcription factors. Fd64A/FoxL1 (asterisk) is most closely related to FoxL1. **B-F''** Fly *FoxL1* is expressed dynamically during embryogenesis. B- F are lateral views; B'-F' are dorsal/ventral views; B''- F'' are FoxL1 antibody staining. B-B''' *FoxL1* is expressed in precursors to the somatic muscles (SMp) and in the hindgut visceral mesoderm (HVMp) at stage 9. C-C'''During stage 12, the mesodermal precursors move with the extending germ band. D-D''' By stage 13, *FoxL1* expression is observed in the hindgut visceral mesoderm and in ten bilateral sets of somatic abdominal muscles. E'E''' *FoxL1* staining begins to wane in the abdominal muscles but appears in the second and third thoracic segments during embryonic stages 14 and 15. F-F''' **G-H** The salivary gland migrates directly adjacent to regions of *FoxL1* expression. **G.** *FoxL1* is expressed in the second and third thoracic segments and in a cluster of cells dorsal to the gland. The salivary gland (outlined) contacts both regions of expression. **H.** FoxL1 is localized to the nuclei, and the salivary gland is positioned between FoxL1-expressing tissues.


A
Mm_FoxE3
Mm_FoxE2
Mm_FoxE1
Mm_FoxD4
Mm_FoxD2
Mm_FoxD3
Mm_FoxD1
Mm_FoxB2
Mm_FoxB1
Dm_fd96Ca
Dm_fd96Cb
Dm_fkh
Mm_FoxA3
Mm_FoxA2
Mm_FoxA1
Dm_fd68a
Mm_FoxK2
Mm_FoxK1
Dm_FoxO
Mm_FoxO6
Mm_FoxO4
Mm_FoxO3
Mm_FoxO1
Dm_CG32006
Mm_FoxM1
Mm_FoxJ1
Mm_FoxJ3
Mm_FoxJ2
Mm_FoxP3
Dm_FoxP
Mm_FoxP4
Mm_FoxP2
Mm_FoxP1
Dm_fd102c
Dm_Crp1
Dm_fd28
Dm_CHES_1-like
Mm_FoxI3
Mm_FoxN2
Mm_FoxR2
Mm_FoxR1
Dm_jumu
Mm_FoxN4
Mm_FoxN1
Mm_FoxH1
Mm_FoxQ1
Dm_bin
Mm_FoxF2
Mm_FoxF1
Dm_fd59a
Dm_fd19b
Mm_FoxL2
Dm_Hlx1
Dm_slp2
Mm_FoxG1
Dm_croc
Mm_FoxC2
Mm_FoxC1
Mm_FoxL1
Dm_fd64a
Mm_FoxS1
Mm_FoxI2
Mm_FoxI3
Mm_FoxI1
Stage 9
Stage 12
Stage 13
Stage 14
Stage 15
B
HVMp
SMp
B'
B''
C
C'
C''
D
SM
D'
D''
E
T2/T3
E'
E''
F
HVM
F'
F''
G
H
αSpec
SAS
Fd64a

in or close to the apical surface, including Stranded at Second (SAS) and Crumbs (Crb). Wild-type salivary glands align relatively parallel to the body wall, along the anterior-posterior axis (**Figure 2A-A'**). Salivary glands from both deficiency lines that remove *FoxL1* showed irregular apical membranes that were often rough and curved (**Figure 2B-C'**), indicating potential migration defects. *Df(3L)BSC369* embryos also had other defects, including irregular visceral mesoderm and gut formation (**Figure 2C**). Each of these lines removes well over twenty genes, indicating that the phenotypes observed are not necessarily due to loss of *FoxL1*. In trans to one another, however, the deficiencies remove only 13 other genes (**Figure 2G**) and still resulted in overt highly penetrant salivary gland defects (**Figure 2D-D',F**). Along with the expression data, these results support a potential role for *FoxL1* in salivary gland migration. Therefore, we decided to more thoroughly investigate the role *FoxL1* plays in embryonic development.

Because null alleles of *FoxL1* did not exist, we created a *FoxL1* null allele by homologous recombination (**Figure 2H**). The *FoxL1* coding region was replaced with the *white+* eye color open reading frame (ORF) by homologous recombination with a clone containing four kb upstream and four kb downstream of the *FoxL1* ORF. Several null alleles were generated, which will heretofore be referred to as $FoxL1^{KO}$. PCR with primer pairs outside the region of recombination and within the *white+* ORF were used to verify insertion of *white+* into the *FoxL1* locus (colored arrows, **Figure 2H**). The candidate $FoxL1^{KO}$ lines yielded PCR products of the expected sizes, whereas wild-type embryos did not, indicating the presence of *white+* in the correct genomic region (Fig 2I). $FoxL1^{KO}$ lines were also tested for transcript and protein expression (**Figure 2J-K''**). Heterozygous $FoxL1^{KO}$ lines had both somatic mesodermal and thoracic *FoxL1* expression (**Figure 2J-J'**), which was absent in their homozygous siblings (**Figure 2K-K'**). Correspondingly, FoxL1 protein expression was also absent in the homozygous $FoxL1^{KO}$ lines (**Figure 2J''-K''**).

Figure 2: Creation of *FoxL1* null allele. A-E' Salivary gland apical membrane staining at stage 15 in *FoxL1* alleles. A- E are lateral views; A'- E' are ventral views. A-A' Wild-type embryos have even and smooth apical membranes. B-C' In the *FoxL1* deficiency lines that remove *FoxL1* and other genes, the apical membranes are rough and irregular. D-D' The deficiencies in trans to one another also result in apical membrane irregularities. E-E' $FoxL1^{KO}$ lines have relatively normal apical membranes. **F.** Quantification of apical membrane irregularities in *FoxL1* mutant alleles. **G.** Genes removed in the Df(3L)BSC368/Df(3L)BSC369 embryos. **H.** Scheme of *FoxL1* knockout to replace the *FoxL1* ORF with *white+* via homologous recombination. Homology arms are approximately four kilobases each. **I.** Agarose gel confirming presence of $white^{+}$ inserted in the *FoxL1* genomic region. Primers are indicated by the colored half arrows in I. **J-K''** $FoxL1^{KO}$ lacks *FoxL1* transcript and protein. J-K' $FoxL1^{KO}$ heterozygotes (as indicated by the balancer chromosome) retain *FoxL1* staining, which is lost in homozygous knockouts. **L-M** Measurements of the distance from the body wall to the salivary gland tip, normalized to embryo width. Measurements were taken for both glands. M. Average of tip to body wall measurements in wild-type and knockout embryos.

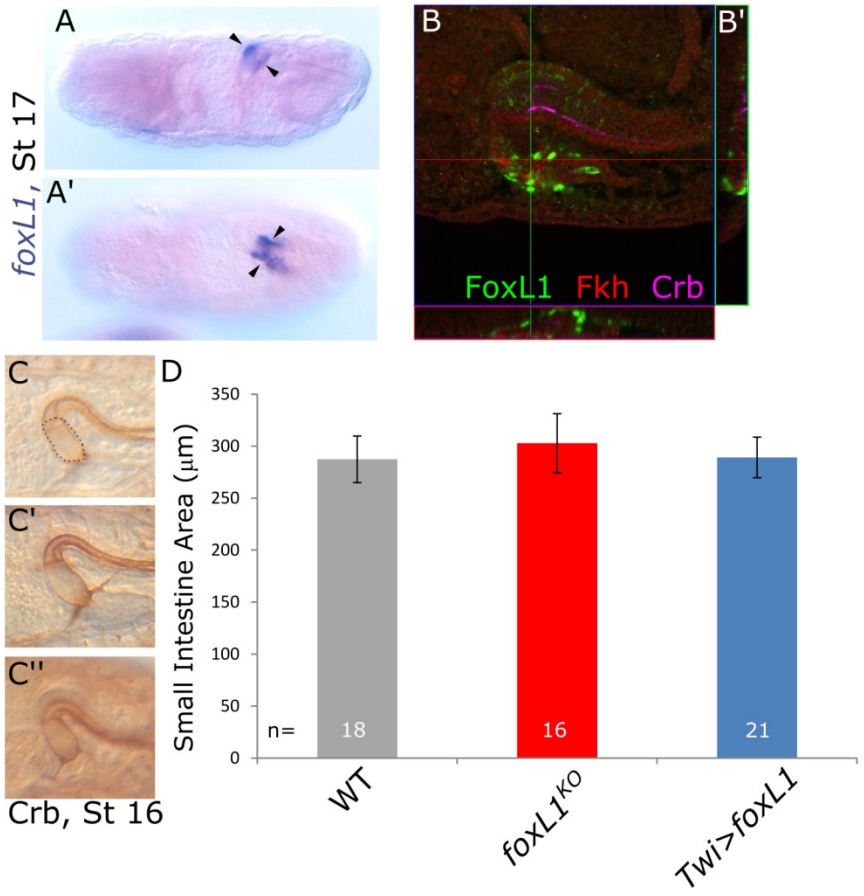

WT
df(3L)BSC368
df(3L)BSC69
df(3L)BSC368/
df(3L)369
foxL1KO
Crb
A
B
C
D
E
A'
B'
C'
D'
E'
F
Percent Irregular Apical Membranes
n=
94
103
104
76
66
WT
df(3L)368
df(3L)368
df(3L)368/
df(3L)369
foxL1KO
G
Genes removed in df(3L)BSC368/df(3L)369
Name
Type of gene
Expression pattern
CG43915
NA
NA
Scrt
Zinc finger DNA binding domain
sensory complex, ventral nerve cord, brain, CNS
CG14985
NA
NA
Mul1
Mitochondria ubiquiton ligase
faint ubiquitous, hindgut
CG11594
cuticle protein
head, hindgut
CR45042
NA
NA
CG11593
Bcl interacting protein
ubiquitous ventral nerve cord
Gr64a
sweet taste receptor
NA
Gr64b
taste receptor
NA
Gr64c
taste receptor
NA
Gr64d
taste receptor
NA
Gr64e
taste receptor
NA
Gr64f
taste receptor
NA
H
3L
foxL1
Mul1
white+
I
Kb
WT
foxL1KO
WT
foxL1KO
fd64aKO/TM6BUL
foxL1KO
St 14
St 15
FoxL1
SG2
J
K
J'
K'
J''
K''
CrebA
L
WT
St 17
L'
foxL1KO
St 17
M
WT
foxL1KO
SG tip to Body Wall/Embryo Width
n=
10
11
14
12
10
12
15
16
17
Stage

Using a single selected null allele, the effects of *FoxL1* loss on salivary gland migration were investigated. The defects seen in *FoxL1*^{KO} were neither as severe nor as penetrant as observed with the deficiency lines (Fig 2E-E',F). Although the apical membrane staining is a sensitive read-out for overall salivary gland morphology, placement and migration, we hypothesized that *FoxL1* could affect gland placement in a different way. Because late embryonic *FoxL1* expression is observed in thoracic body wall segments (**Figure 1H**), glands were examined from a ventral view and assayed for their position relative to the embryonic body wall (**Figure 2L-L'**). The distance from the body wall to the salivary gland tip was measured for both glands and divided by the width of the embryo to generate a normalized ratio of the distance from gland tip to body wall. No significant difference was observed between wild-type embryos and *FoxL1*^{KO} embryos (**Figure 2M**). Therefore, by the assays used, loss of *FoxL1*^{KO} had only very minor effects on salivary gland placement.

Hindgut morphology is normal in FoxL1 mutants

FoxL1 is also expressed at or near the hindgut (**Figure 3A, A'**). Co-staining with FoxL1, Crb (a marker for the apical surface of epithelial cells), and Fkh (a transcription factor expressed in the hindgut endoderm; **Figure 3B**) reveals that FoxL1 is not expressed in the hindgut endoderm, but is instead expressed in cells that cover the endoderm. These cells are most likely the hindgut visceral mesoderm, although staining in a *heartless* or *wingless* mutant background (which lack the HVM) would be necessary to confirm this localization (San Martin and Bate, 2001). To ask if *FoxL1* is required for hindgut morphology, wild-type and *FoxL1*^{KO} homozygous embryos were stained with Crb and the cross-sectional area of the small intestine, the portion of the hindgut outlined in Figure 3C, was measured. No significant difference in the small intestine area was

Figure 3: Loss and overexpression of FoxL1 does not affect hindgut morphology. A-A' Lateral (A) and ventral (A') views showing *FoxL1* transcript around the hindgut at Stage 17. **B-B'** FoxL1 staining surrounds the hingut ectoderm (Fkh) and the hindgut tube (Crb). **C-C''** Small intestine of the hindgut in *FoxL1* alleles. C. Wild type embryos. Small intestine is outlined. C' $FoxL1^{KO}$. C'' Twi>*FoxL1* mis-expression. **D.** Quantification of the area of the small intestine in the different genotypes.


A
A'
foxL1, St 17
B
B'
FoxL1 Fkh Crb
C
C'
C''
Crb, St 16
D
Small Intestine Area (µm)
350
300
250
200
150
100
50
0
n=
18
16
21
WT
foxL1KO
Twi>foxL1

observed in the *FoxL1* loss-of-function background, suggesting that *FoxL1* does not overtly affect the development or morphology of this portion of hindgut.

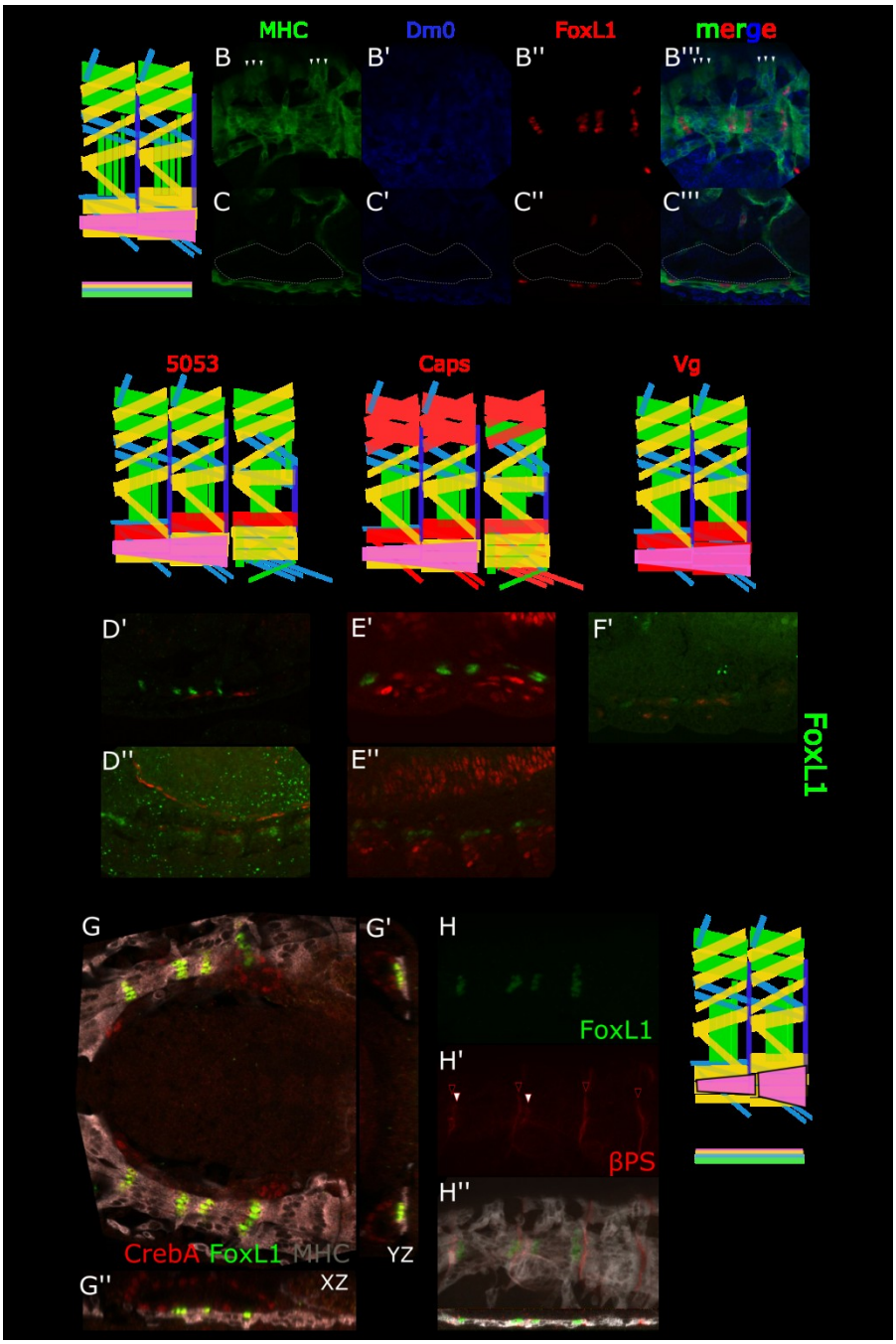
The effect of *FoxL1* over-expression was also examined. Since a full-length cDNA of *FoxL1* was not available, one was generated using genomic DNA as a template. First, the genomic region of the *FoxL1* ORF was subcloned into a TOPO vector. Two consecutive rounds of deletion PCR were then used to remove each intron (**Figure 6A**). Deletion of the introns was confirmed using diagnostic digests, which revealed a size shift and the loss of a restriction site (**Figure 6B**). The FoxL1 cDNA was further confirmed by sequencing the entire ORF. This clone was subsequently used to create the UAS-*FoxL1* line. Overexpression of *FoxL1* in the mesoderm using *twist*-Gal4 also did not affect hindgut morphology (**Figure 3C**).

FoxL1 is expressed in muscle VIS5/muscle 33

FoxL1 is expressed in a subset of ventrally-positioned somatic muscles in both the thoracic and abdominal segments of the embryo. The musculature of the *Drosophila* embryo is organized into approximately thirty muscle fibers in each segment (**Figure 4A**). Slight differences exist between the abdominal and thoracic segments, most notably near the mouthparts and anus. In the first through third thoracic segments (T1-T3), an extra muscle is found that is proposed to span all three segments and attach to the mouthparts (**Figure 4A**, pink trapezoid). This muscle is known as muscle 33 or Ventral Intersegmental 5 (VIS5), and it is the most internally localized muscle. The other muscles, which are found in both the thoracic and abdominal segments, are external and adjacent to VIS5 (yellow layer) or external and in direct contact with the body wall (green layer).

To verify that FoxL1 is expressed in muscles, embryos were costained with FoxL1, MHC (a muscle marker), and Dm0 (a nuclear lamina marker; **Figure 4B-C'''**). FoxL1 expression aligns with

Figure 4: Identification of VIS5 as the thoracic muscle expressing FoxL1. **A.** Schematic of thoracic segment two and three muscles. Top: Lateral view of muscle organization. Green muscles are the most external, followed by blue, yellow, and pink. Bottom: ventral view of muscle organization. Lateral transverse (LT) muscles are outlined. **B-C'''** FoxL1 is expressed in the ventral muscles. B-B''' Lateral view of musculature and FoxL1 staining. Arrowheads indicate the LT muscles. C-C''' Ventral views of Fd4a and musculature. FoxL1 is expressed in the most internal muscles, which are directly next to the outlined salivary gland. **D-D''** FoxL1 is not expressed in VLM1. **D.** 5053-Gal4 drives nuclear LacZ in VLM1 (red). D'-D'' Costaining of LacZ and FoxL1 in the thoracic and abdominal muscles. **E-E''** FoxL1 is not expressed in VLM1 or the ventral oblique muscles, marked by a Caps-LacZ reporter. **E-E''** Costaining of LacZ and FoxL1 in the thoracic and abdominal muscles. **F-F'** FoxL1 is not expressed in VLM1-4, marked with Vg (red). **G-G''** FoxL1 is expressed in VIS5. CrebA (Salivary gland nuclei), FoxL1, and MHC (muscle) staining shows that FoxL1 is expressed in muscles directly next to the migrating salivary gland (G''), the most internal muscles (G'). **H-H''** VIS5 has muscle attachments sites offset from VLM attachment sites. **H'** FoxL1 staining. **H'** βPS staining showing muscle attachment sites. VLM attachments are marked with open arrowheads. Additional attachments sites are marked with white arrowheads. **H''** Merge, showing the muscle attachment sites that are offset from the VLM attachment sites correspond to FoxL1 staining. Musculature is shown with MHC staining in gray. **I.** VIS5 (pink) is an individual muscle in each of the thoracic segments.


MHC
Dm0
FoxL1
merge
B ,,,
B'''
B'
B''
B''',,,
C
C'
C''
C'''
5053
Caps
Vg
D'
E'
F'
D''
E''
FoxL1
G
G'
H
FoxL1
H'
βPS
H''
CrebA FoxL1 MHC
YZ
G''
XZ

the negative space in MHC staining (Fig 4B) and corresponds to predicted nuclear staining (**Figure 4B'**). FoxL1 appears to be expressed in or near the ventral longitudinal muscles (VLM), which are ventral to the vertically aligned lateral transverse muscles (**Figure 4A**). This assignment is further supported by examination of ventral views of the muscles, which border the salivary gland, a ventrally-positioned structure (**Figure 4C-C'''**).

In the thoracic segments, there are three individual VLMs, whereas in the abdominal segments there are four. To determine if FoxL1 is indeed expressed in these muscles, costaining with a variety of specific VLM markers was performed. 5053-Gal4 drives nuclear lacZ staining in only VLM1 (**Figure 4D**). No overlap between FoxL1 and LacZ was observed in either the thoracic or abdominal segments (**Figure 4D'-D''**). A Capricious (Caps) reporter was used to mark VLM1 along with a subset of the ventral oblique muscles (VO) and the dorsal oblique and anterior (DO and DA) muscles. Again, no overlap was observed in T2/T3 or with the abdominal muscles expressing FoxL1. Finally, Vestigial was used to mark all of the VLM muscles (**Figure 4F**). Due to technical constraints, staining of Vg in the abdominal segments with FoxL1 staining was not obtained. In T2/T3, however, both antibodies worked, and no overlap between the markers was observed (**Figure 4F'**). We conclude that FoxL1 is not expressed in the VLM muscles.

Based on staining with the available muscle markers described above, it appears that FoxL1 is expressed in VIS5. As noted before, FoxL1 is expressed in the musculature near the salivary gland (**Figure 4G**). Analysis in the Z-axis reveals that FoxL1 is in muscle cells directly contacting the salivary gland (**Figure 4G'**), the most internal muscle group (**Figure 4G''**). In T2/T3, this corresponds to VIS5. However, VIS5 was previously described as one large muscle spanning all three thoracic segments. Nuclei in Drosophila muscles move to the edges of the muscles (Folker et al 2012). Given the placement of FoxL1-expressing nuclei, two possibilities arose: (1) FoxL1 is not expressed in the VLMs or VIS5, or (2) VIS5 is not one large muscle. To test the second

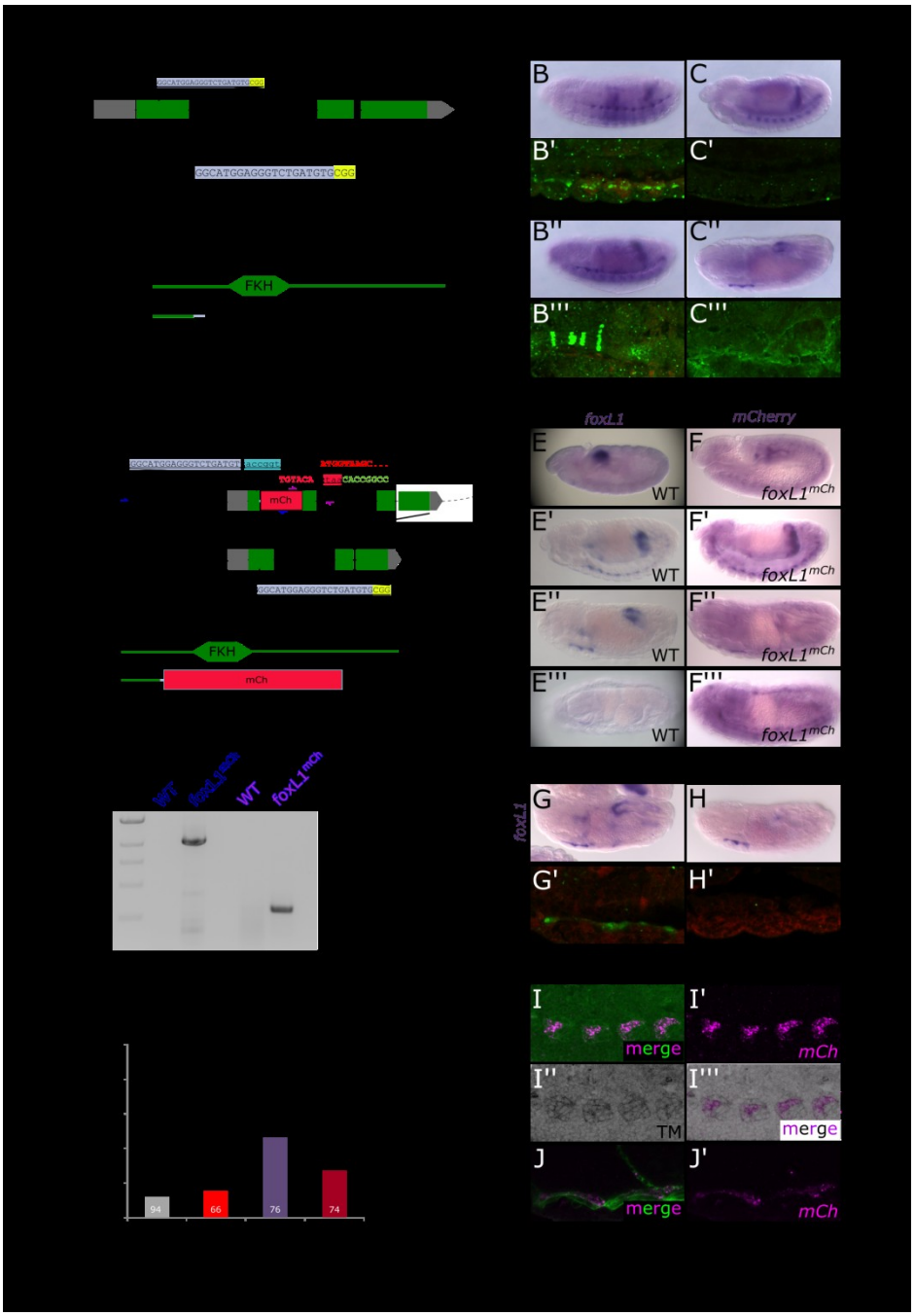
hypothesis, embryos were stained with FoxL1, MHC, and βPS, which marks muscle attachment sites (**Figure 4H-H''**). Muscle attachment sites in the VLM region correspond to the abdominal segment boundaries. Indeed, muscle attachment sites are present at the segment boundaries in the thorax and are found slightly offset from the boundaries of other thoracic muscles (**Figure 4H'**, arrowheads), indicating separate attachment sites for the most internal muscles (XZ view). Therefore, we conclude that that an individual VIS5 muscle is present in each of the three thoracic segments, and that FoxL1 is expressed in VIS5 of the second and third thoracic segments (**Figure 4I**).

Creating a new, tagged, allele of FoxL1 to assay for muscle phenotypes

Since we had no good markers to evaluate *FoxL1* function in the muscles that normally express *FoxL1,* we decided to replace the coding region of *FoxL1* with a fluorescent tag using the CRISPR/Cas9 system (**Figure 5A, D**). We first used the system to generate small nucleotide deletions in the *FoxL1* gene. A PAM site was selected that targeted the second exon of *FoxL1*, which corresponds to a region approximately 150 nucleotides downstream from the start codon (**Figure 5A**). Several new *FoxL1* alleles were generated (**Figure 5A'**), with a variety of small deletions near the PAM site. The line with the cleanest disruption of *FoxL1*, $FoxL1^{10.5}$ deleted two base pairs (**Figure 5A'**), resulting in a truncation of FoxL1 (**Figure 5A''**). The first 54 amino acids of FoxL1 are in frame, followed by an out-of-frame region of 11 residues, followed by a nonsense mutation. The resulting protein is predicted to be 65 residues and ends before the Forkhead box DNA binding domain.

The $FoxL1^{10.5}$ allele was tested for transcript and protein expression. Compared to heterozygous siblings, $FoxL1^{10.5}$ had similar levels of *FoxL1* transcript (**Figure 5B-C''**). Importantly,

Figure 5: Creation of *FoxL1* alleles using CRISPR/Cas9. A-A'' Creation of null alleles using CRISPR. A. *FoxL1* genomic region. The PAM site and the sequence of the CRISPR region are shown in yellow and gray, respectively. The PAM site is located in the first coding exon. **A'** Three alleles created by the CRISPR site result in small deletions near the PAM site. **A''** *FoxL1*$^{10.5}$ results in a protein truncation before the Fkh DNA binding domain. **B-C'''** *FoxL1*$^{10.5}$ is protein null. **B-C''** Stage 14 embryos; **B-C** *FoxL1*$^{10.5}$ heterozygotes (**B**) and homozygotes (**C**) have *FoxL1* transcript. Only the hetrozygotes (**B'**) have protein. Stage 16 embryos are shown in **B''-C'''**. **D.** *FoxL1* CRISPR creating a knock-in of mCherry. CRISPR region is shown in gray; PAM in yellow. In the homologous recombination construct, two kb of *FoxL1* directly flanking the PAM site were cloned into p-mCH-C1 flanking the mCherry ORF. The PAM site was mutated to prevent (re)targeting (*), residues were added to keep mCherry in frame, and a stop codon was added at the end of mCherry (red). **D'** *FoxL1*^{Mch} creates a fusion protein. **D''** PCR with primers indicated in D (half arrow) confirming insertion of mCherry. **E-F'''** The *FoxL1*^{Mch} expression pattern reflects the spatiotemporal profile of the endogenous *FoxL1* transcript. **G-H'** *FoxL1*^{Mch} has *FoxL1* transcript but not protein expression. **I-J'** FISH of *FoxL1*^{Mch} shows *mCherry* expression in the abdominal segments muscles and VIS5. Musculature is shown by Tropomyosin staining (**I''**; green in **I** and **J**).


FKH
mCh
WT
foxL1mCh
WT
foxL1mCh
94
66
76
74
B
C
B'
C'
B''
C''
B'''
C'''
foxL1
mCherry
E
F
WT
foxL1mCh
E'
F'
E''
F''
E'''
F'''
G
H
G'
H'
I
merge
I'
mCh
I''
TM
I'''
merge
J
merge
J'
mCh

protein expression was completely absent in *FoxL1*$^{10.5}$ in both the abdominal and thoracic segments (**Figure 5B'-C''**), indicating that *FoxL1*$^{10.5}$ is protein null.

The same PAM site was used for homologous recombination. The template for homologous recombination was the p-mCh-C1 vector, which had two two kb homology arms of the *FoxL1* genomic region flanking the coding region for the mCherry fluorescent protein. The region of homology began directly on either side of the PAM site. Additionally, the PAM site in the vector was altered so that (1) it would no longer be a target for the CRISPR construct, and (2) would keep the mCherry gene in frame with the start of *FoxL1*. A stop codon was also inserted at the end of the mCherry ORF. Our goal was for the recombined region to both report on *FoxL1* expression and create a null allele. Individual lines were isolated that all had mCherry inserted in frame with *FoxL1*, resulting in a fusion protein that included the 80 N-terminal residues of FoxL1, a short linker region, and full length mCherry (**Figure 5D'**). PCR with primers within mCherry and in the homology arms was used to confirm the insertion of mCherry (**Figure 5D**, arrows; **Figure 5D''**). An in situ probe to *mCherry* was also tested on these lines, and it accurately reports on the *FoxL1* expression pattern at each developmental stage (**Figure5E-F'''**). Like *FoxL1*$^{10.5}$, *FoxL1*^{Mch} contains low levels of *FoxL1* transcript, but is protein null (**Figure 5G-H'**).

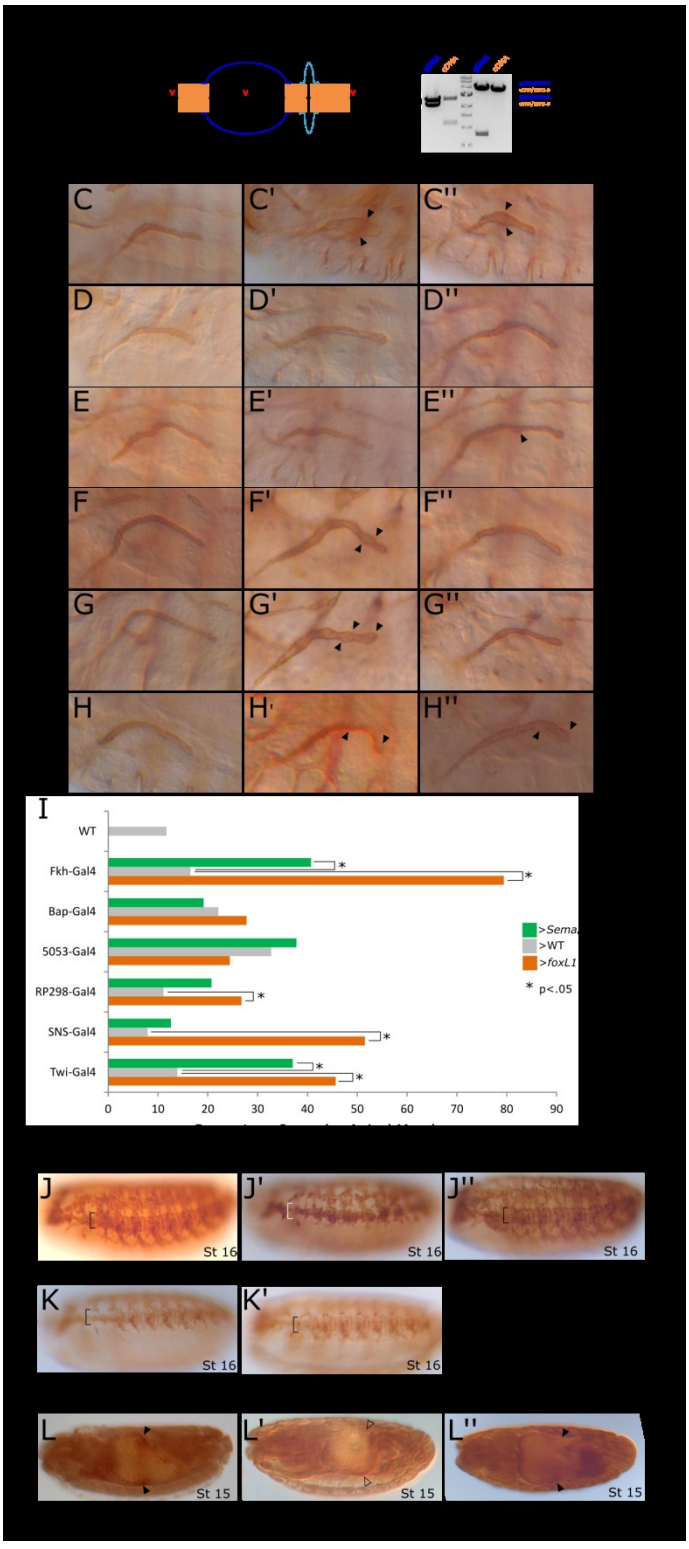
Salivary gland migration was examined in both *FoxL1*$^{10.5}$ and *FoxL1*^{Mch} (**Figure 5K**). The same types of minor apical membrane irregularities were seen with the new alleles as were seen with the *FoxL1*^{KO}allele, but at a slightly higher penetrance. Because of the differences in penetrance, future analysis will be conducted with transheterozygotes of *FoxL1*$^{10.5}$ and *FoxL1*^{KO}. *FoxL1*^{Mch} was used to examine muscle morphology. Two separate mCherry antibodies were tested on this line, and neither generated a positive signal, and no mCherry signal was observed in live embryos (data not shown). This suggests either that the antibodies are not compatible with the fixative used, that the fusion protein is being degraded or that the small N-terminal region of

FoxL1 is somehow disrupting mCherry 3D structure. While this issue is being resolved, I performed FISH plus antibody staining (**Figure 5I-J'**). At stage 13, *mCherry* is observed in one muscle per segment in the abdominal segments (**Figure 5I''**). At this stage the muscles have not yet formed their final pattern. Shortly after stage 13, the staining of *mCherry* (or FoxL1) in these muscles disappears. Closer examination will be necessary to determine if this muscle dies or just loses FoxL1/*mCherry* expression. At later stages, *mCherry* is observed in VIS5 (**Figure 5J-J'**). Additional experiments must be completed to fully understand the shape and placement of VIS5 in these embryos, but it must be noted that in $FoxL1^{Mch}$, the expression of *mCherry* indicates that VIS5 is present in both T2 and T3. If there are defects in this muscle, they may be subtle and will require detailed examination.

FoxL1 over-expression disrupts the morphology of multiple tissues

The absence of any overt effects with loss of *FoxL1* on salivary gland positioning despite the proximity of the muscle staining suggested that perhaps *FoxL1* regulates expression of redundant signals used for salivary gland navigation. To test this possibility, *FoxL1* was expressed using a variety of Gal4 lines and salivary gland migration was examined. Mis-expression of *FoxL1* in muscle 12 (*5053-Gal4*; **Figure 6E-E'**) or in the visceral mesoderm (*bap-Gal4*; **Figure 6D-D'**) had very little effect on salivary gland placement or apical morphology (**Fig 6I**). On the other hand, mis-expression of *FoxL1* in all fusion competent myoblasts (*SNS-Gal4*; **Figure G-G'**), in muscle founder cells (*RP298-Gal4*; **Figure 6F-F'**), or in all muscle types (*Twi-Gal4*; **Figure 6H-H'**) all caused significant irregularities in the salivary gland apical membrane, consistent with migration defects. These drivers are expressed early and result in persistent *FoxL1* expression (data not shown).

Figure 6: *FoxL1* mis-expression results in salivary gland irregularities. A. Creation of *FoxL1* cDNA through two rounds of deletion PCR. B. PCR confirming deletion of *FoxL1* introns. **C-H''** Apical membrane staining of wild-type (C - H), *FoxL1* mis-expression (C' - H') or *sema2aTM-GFP* (C'' - H''). Gal4 drivers are indicated at the left margin. **I.** Quantification of apical membrane irregularities. **J-K'** Muscle staining of embryos mis-expressing *FoxL1*. J'. Mis-expression of *FoxL1* in all muscle by Twi-Gal4 causes severe defects in muscle organization (white bracket) compared to wild-type embryos (black bracket; J). Mis-expression of *sema2a-TM-GFP* does not cause the same defects (bracket; J''), nor does mis-expressing *FoxL1* with other drivers (brackets; K-K'). **L-L''** Mis-expression of *FoxL1* causes gut defects and loss of Labial staining.

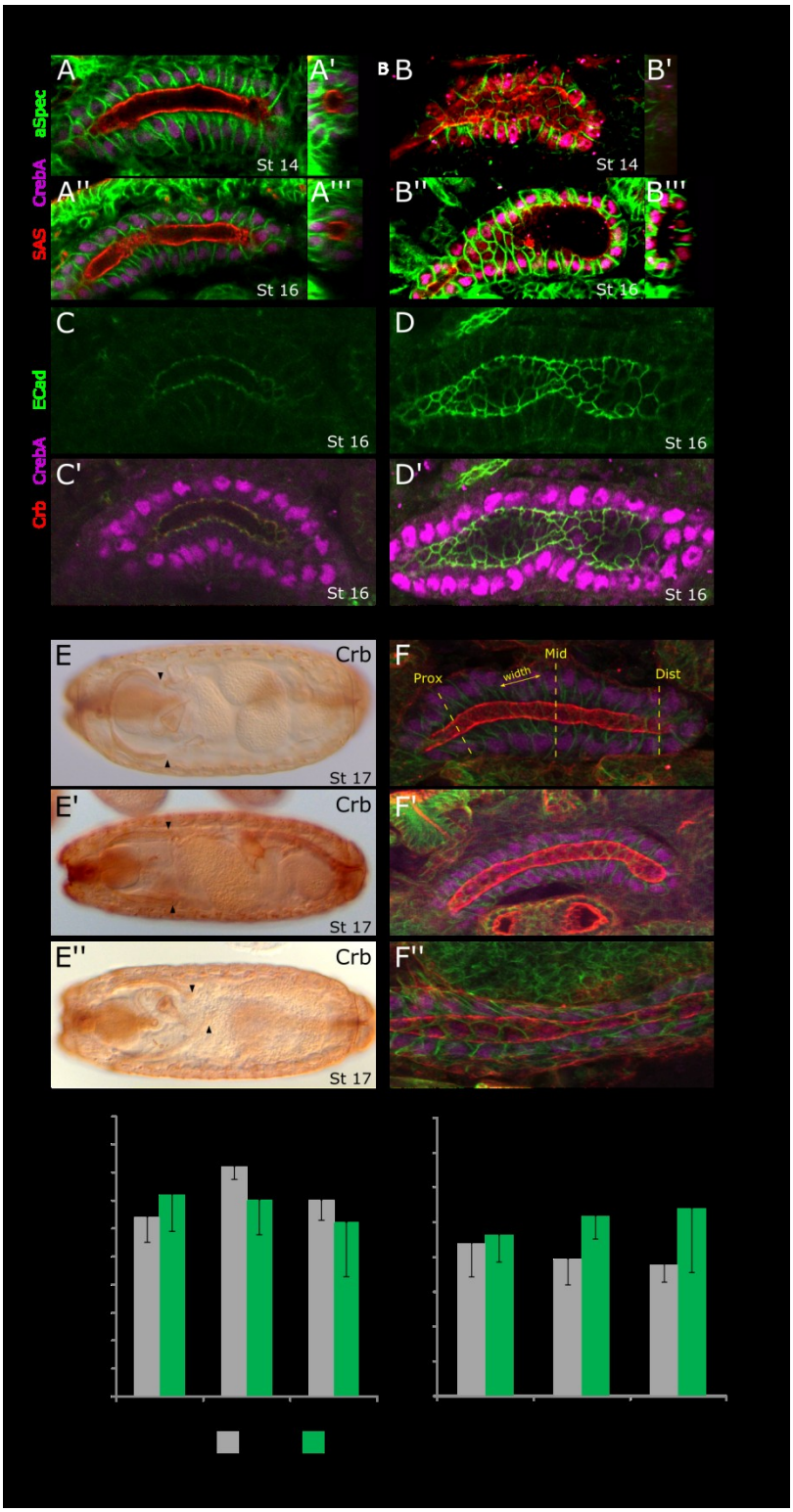

C
C'
C''
D
D'
D''
E
E'
E''
F
F'
F''
G
G'
G''
H
H'
H''
I
WT
Fkh-Gal4
Bap-Gal4
5053-Gal4
RP298-Gal4
SNS-Gal4
Twi-Gal4
>Sema
>WT
>foxL1
* p<.05
0
10
20
30
40
50
60
70
80
90
J
J'
J''
St 16
K
K'
St 16
L
L'
L''
St 15

Mis-expression of *FoxL1* in the fusion competent myoblasts, founder cells, or pan-musculature could be driving the transcription of a signal that affects the salivary gland, or it could be driving the transcription of a signal that affects muscle formation itself. To examine the musculature in these different overexpression situations, embryos were stained with an antibody to the Myosin Heavy Chain (MHC; **Figure 6J-K'**). By stage 16, the VLM muscles are organized into four parallel fibers that run along the anterior-posterior length of the embryo (bracket; **Figure 6J**). In Twi-Gal4>UAS-*FoxL1* embryos, the entire musculature is severely disorganized (**Figure 6J'**). This could suggest that the salivary gland defects are due to the impairment of the scaffold provided by the musculature. Since, however, muscle architecture is not affected in RP298-Gal4>UAS-*FoxL1* or SNS-Gal4>UAS-*FoxL1* embryos (**Figure 6K, K'**), the salivary gland defects seen in these embryos may not due to alterations of the scaffolding alone, and, instead, may be due to ectopic expression of a FoxL1-dependent signal(s) that is redirecting migration of the salivary gland .

Notably, while scoring the Twi-Gal4>UAS-*FoxL1* embryos, we observed that these embryos often had malformed guts with holes or lesions in the various compartments **(Figure 6L-L'')**. Wild-type and Twi-Gal4>UAS-*FoxL1* embryos were stained with Labial, which marks the cuprophilic cells that populate the second gut compartment (Tremml and Bienz, 1992). At stage 15, these cells are located near the second gut constriction (**Figure 6L**). Many Twi-Gal4>UAS-*FoxL1* embryos are Labial negative, indicating that the second gut compartment is not being correctly specified (**Figure 6L'**). Whether this phenotype is due to impaired muscle function or impaired signaling has yet to be determined.

Mis-expression of *FoxL1* in the salivary gland caused major gland defects, most notably an inflated apical lumen (**Figure 6C'**). This phenotype warranted a closer examination. Whereas nuclear positioning and apical/lateral polarity (**Figure 7A-B'''**) were not affected with *FoxL1*

Figure 7: Mis-expression of *FoxL1* in the salivary gland and in all muscles cause changes in salivary gland organization. A-B’’’ Confocal imaging of wild-type (**A-A’’’**) and *FoxL1* mis-expression (**B-B’’’**) in the salivary gland. The shape of the apical membrane and lumen is deformed in *fkh>FoxL1*, as revealed by the YZ sections (A’, A’’’; B’, B’’’). **C-D’’** ECadherin staining of WT (C-C’) and Fkh>FoxL1 salivary glands. **E-F’’** Apical membrane length in wild-type, twi-Gal4>UAS-FoxL1, and twi-Gal4>UAS-sema2aTM-GFP embryos. The ends of the glands are marked with arrowheads. Mis-expression of *FoxL1* by Twi-Gal4 causes abnormally long apical membranes (E’’, F’’), a phenotype not observed with *sema2a-TM-GFP* mis-expression (C’, D’). **G-H** Quantification of cell parameters indicated in (D) for WT and *FoxL1* mis-expressing embryos.


A
A'
A''
A'''
B
B'
B''
B'''
aSpec
CrebA
SAS
St 14
St 16
C
C'
D
D'
ECad
Crb
E
E'
E''
F
F'
F''
Crb
St 17
Prox
width
Mid
Dist

misexpression, the YZ plane reveals that the lumen of the gland is very narrow at stage 14 (**Figure 7B'**), but then inflates to an overly large size by stage 16 (**Figure 7B'''**). This is in contrast to control salivary gland lumens, which maintain an even size from stage 14 to stage 16 (**Figure 7A', A'''**). The cells in *fkh-Gal4>UAS-FoxL1* also appear more cuboidal than the wild-type columnar gland cells (**Figure 7A'', B''**) and more cells surround the lumen (**Figure 7A''', B'''**) Interestingly, E-Cadherin levels appear to be upregulated in *FoxL1* expressing salivary glands (**Figure 7C-D'**).

The gland defects when *FoxL1* was mis-expressed in all muscles were examined more closely. The glands become over-elongated and often extend into the region of the gut compartments in *Twi-Gal4>UAS-FoxL1* embryos (**Figure 7C, C''**). Both the number of nuclei around these abnormally long glands and the width of the cells of the gland were quantified (**Figure 7D-D''**). Whereas neither data set is significantly different by standard deviation measures, two trends were observed. One, the number of nuclei around the tube in *Twi-Gal4>UAS-FoxL1* glands was constant from the proximal to the distal end of the tube (**Figure 7E**). Two, the cells in Twi-Gal4>UAS-*FoxL1* glands are slightly elongated (along the long axis of the tube) compared to control glands (**Figure 7F**). These phenotypes suggest that modulation of *FoxL1* levels in muscles affects cell re-arrangement and cell morphology within the salivary gland, perhaps by providing an attractive signal towards which the gland extends.

Sema2a functions downstream of FoxL1

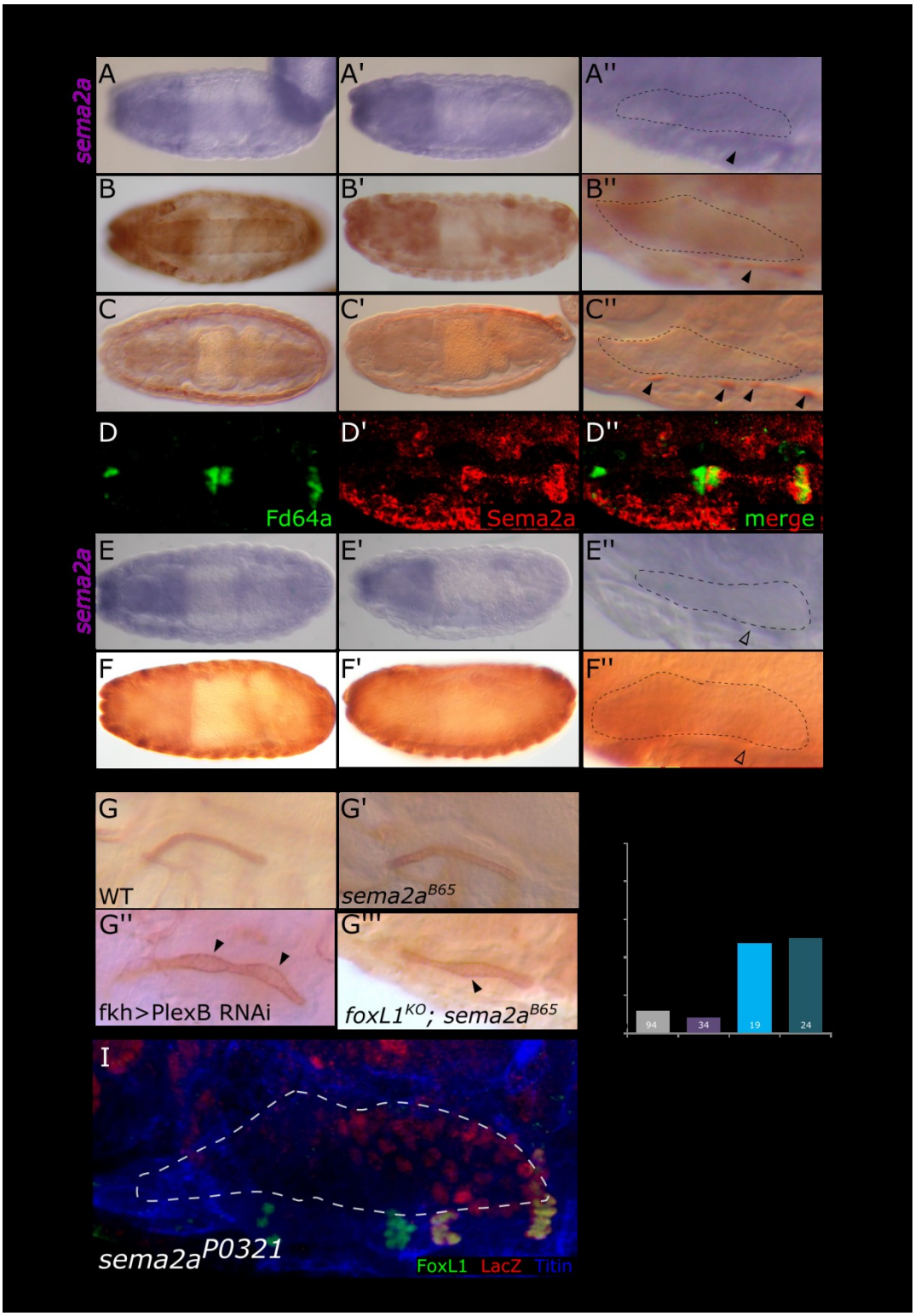
The non-autonomous effect of *FoxL1* overexpression on salivary gland morphology suggests that it controls expression of genes that encode secreted signals. One excellent candidate target of FoxL1 is *sema2a*, which encodes a member of the Semaphorin family of signaling molecules. Sema2a is known to function as a repellent for axon guidance (Ayoob et al 2006). *sema2a* is

expressed in T3 (**Figure 8A-A''**), in a region that overlaps *FoxL1* expression. Sema2a protein is observed in the same area (**Figure 8B-B''**); and co-staining with FoxL1 and Sema2a revealed that secreted Sema2a is expressed in VIS5 of T3, where FoxL1 is also expressed (**Figure 8D-D''**). To ask if *sema2a* is a target of FoxL1, both *sema2a* RNA and Sema2a protein were examined in the $FoxL1^{KO}$ line. Indeed, transcript and protein were missing in the VIS5 muscle of T3 in the *FoxL1* null embryos, indicating that *sema2a* is indeed a target of FoxL1 (**Figure 8E-F''**).

As with *FoxL1*, both loss-of-function and misexpression of *sema2a* were tested for salivary gland defects. The $sema2a^{B65}$ mutants showed no overt defects in the salivary gland (**Figure 8G-G'**). However, as with *FoxL1*, over-expression of *sema2a* caused significant salivary gland defects (**Figure 6C''-H''**). Mis-expression using RP298 (**Figure 6F''**), SNS (**Figure 6G''**), or Twi (**Figure 6H''**)-Gal4 caused significant irregularities in the apical membrane of the salivary gland (**Figure 6I**). Defects were also observed by over-expressing *sema2a* in the salivary gland. However, mis-expression of *sema2a* in all muscles did not cause the same defects as mis-expression of *FoxL1* (**Figure 6K''; Fig 7C', 7D'**). One reason for the discrepancy in the phenotypes may be that FoxL1 regulates other signals in addition to Sema2a (**Figure 8I**).

Interestingly, both Sema2a and the receptor for Sema2a, PlexB, are expressed in the salivary gland (Melissa Vining, unpublished); Sema2a is expressed in the distal half of the salivary gland and PlexB is expressed in all salivary gland cells. Perhaps the salivary gland expression of Sema2a and/or the Sema2a receptor somehow account for the normal salivary gland appearance in *sema2a* and *FoxL1* null alleles. To explore this possibility, we examined the salivary gland apical membrane in *plexB* knockdown embryos and in embryos lacking both *sema2a* and *FoxL1*. Indeed, these embryos showed an increase in salivary gland apical membrane irregularities (**Figure 8G''-H**). These defects suggest that *FoxL1*, *Sema2a* and otherdownstream signals play a role in guiding salivary gland migration.

Figure 8: *sema2a* is a target of *FoxL1*. A-B'' Sema2a is expressed in T3. A-A'' *sema2a* transcript is seen in the third thoracic segment. **A**,**B** lateral; **A'**, **B'** ventral, **A'', B''** higher magnification showing the *sema2a* expression is directly next to the salivary gland. **B-B''** Sema2a antibody staining reflects the transcript expression. **C-C''** FoxL1 staining showing expression in VIS5 of T2 and T3. **D-D''** Sema2a is expressed in VIS5 of T3. **E-F''** Sema2a staining is lost in $FoxL1^{KO}$. **E-E''** *sema2a* transcript is shown; **F-F''** Sema2a antibody staining is shown. **G-G'''** Salivary gland staining of WT (G), *sema2a* null allele (G'), RNAi knockdown of *plexB* (G''), and *sema2a, FoxL1* double null allele (G'''). **H.** Quantification of apical membrane irregularities. **I.** Sema2a is also expressed in the salivary gland based on in situ expression in earlier embryos (not shown) and on nuclear lacZ expression from $sema2a^{P0321}$, a *sema2a* reporter (red).


sema2a
A
A'
A''
B
B'
B''
C
C'
C''
D
D'
D''
Fd64a
Sema2a
merge
sema2a
E
E'
E''
F
F'
F''
G
WT
G'
sema2a^B65
G''
fkh>PlexB RNAi
G'''
foxL1^KO; sema2a^B65
94
34
19
24
I
sema2a^P0321
FoxL1 LacZ Titin

DISCUSSION

FoxL1 is one of the 19 members of the Fox family of transcription factors encoded in *Drosophila*. *FoxL1* is dynamically expressed in different subsets of somatic mesodermal cells throughout embryogenesis. At early stages, *FoxL1* is expressed in mesodermal precursors that migrate anteriorly during germ-band extension. *FoxL1* is expressed in ten bilateral clusters of cells In the abdominal segments during embryonic stage 13. The abdominal clusters of FoxL1-expressing cells remain unidentified, since staining did not colocalize with any of the existing somatic muscle markers available. This cluster of cells could be precursors to a muscle homologous to VIS5 in the abdominal segments. A muscle analogous to VIS5 is made in the first abdominal segment, but it is absent in the other segments. One hypothesis is that the precursor is formed in the other abdominal segments, but later dies, potentially due to lack of a cofactor that is present in the thoracic and first abdominal segment. Our attempts to co-stain these clusters with antibodies that label dying cells were unsuccessful due to the different fixation protocols required for each antibody. Cell death is not known to occur in muscles during normal embryogenesis (Bate 1990), so another possibility is that these clusters of cells are fusion competent muscle cells and fuse into a different muscle fiber.

Drosophila muscles form by fusion of founder cells (FCs), which specify individual muscles, and fusion competent myoblasts (FCMs), which are a more naïve cell type that then adopts the identity of the FC they fuse with (Baylies et al 1998). Fusion occurs in two rounds during development. The first round of fusion happens around stage 12-13, while the second round, which contains the majority of fusion events, occurs during stages 14-15 (Beckett et al 2007). FoxL1 may mark FCMs in the abdominal segment that fuse during the first round of fusion. Co-staining of FoxL1 with an FCM marker such as Lameduck would address this hypothesis. *mCherry* expression in *FoxL1*^{mCh} reveals that the FoxL1-positive muscles appear to form a fiber

(**Figure 6I'''**), which would argue against FoxL1 being expressed in a naïve cluster of FCM cells. FCs are specified through the combined interaction of transcription factors (Baylies et al 1998). Different combinations of transcription factors specify different FCs, while loss of one type of transcription factor can cause an FC to form a different type of muscle (Frasch 1999). FoxL1 expression may aid in specification of a specific type of muscle in the abdominal segments, whereas loss of FoxL1 may cause this muscle to be specified differently. More careful analysis of the abdominal muscles and the counting of nuclei may be necessary to determine the role of FoxL1 in these muscle cells. Meanwhile, the identify and fate of the ten bilateral clusters of *FoxL1* positive abdominal muscle cells remains elusive.

FoxL1 is also expressed in VIS5 muscle in the somatic muscles of the thoracic segments. Because it is not expressed until later in development in VIS5, *FoxL1* is unlikely to be acting as a FC marker for this muscle. However, because the internal muscles are the last to fuse with FCMs (Richardson et al 2008), we cannot rule out this possibility. Our analysis has shown that VIS5 is an individual muscle in the three thoracic segments, not one large muscle spanning three segments as previously suggested (Bate, 1990). This finding was revealed by two observations. One, *FoxL1* RNA does not extend into the first thoracic segment, suggesting a barrier between T1 and T2, and two, the presence of muscle attachment sites for the most internal muscles in T2 and T3. Staining with FoxL1 showed that it is expressed in the most internal ventral muscles of the thoracic segments, which correspond to VIS5. The only known role of VIS5 is that it expresses Sema2a, which affects muscle innervations, but only when misexpressed (discussed more below; Kolodkin et al 1993). Very few reagents exist to study the VIS5. Starvin (Stv) was reported to be expressed in VIS5 (Coulson et al 2005), as well as the tendon cells and the esophagus. Loss of *stv* causes retraction or bunching of the muscles in the thoracic segments, and delayed growth, likely due to feeding difficulties. The authors do not show co-staining of Stv

with any other VIS5 markers, and I was unable to replicate the staining seen with the Stv antibody with a *stv* in situ probe (data not shown). No other VIS5 specific markers exist. To have VIS5 specific marker, we used CRISPR/Cas9 to knock mCherry in the *FoxL1* ORF. This line accurately mimics *FoxL1* expression with *mCherry*, and provides a way to further investigate VIS5 and the other tissues that express *FoxL1*. Unfortunately, mCherry protein could not be detected with the two antibodies we tested.

The expression of *FoxL1* in the thoracic segments is interesting because the salivary gland migrates directly between cells that express *FoxL1*. The salivary gland contacts VIS5 on its ventrolateral surface beginning in stage 14. The migrating salivary gland also contacts a *FoxL1* positive cluster of muscles in the head during stage 13. Due to this expression pattern, we hypothesized that *FoxL1* influenced gland positioning during migration by providing an attractive signal. Staining of the salivary gland showed no major defects in gland morphology or placement with either *FoxL1* null allele. However, mis-expression of *FoxL1* in the salivary gland caused an increase in E-cad levels and an inflation of the gland lumen. This increase of E-cad may enhance cell-cell adhesion in the salivary gland, thus preventing the cell rearrangements that normally contribute to the correct lumen size. Interestingly, pan-mesodermal expression of *FoxL1* causes the musculature to clump. One hypothesis is that over-expression of *FoxL1* is causing increased E-Cad levels in the muscles, causing them to adhere to one another instead of spreading into fibers. Alternatively, cells expressing *FoxL1* may simply be highly attracted to one another and this increase in cell attraction may result in the increased levels of E-Cad.

Myotubes in Drosophila get to their correct location using guidance cues such as Robo-Slit and Wnt-Frizzled (Kramer et al 2001; Ghazi et al 2003). These same signaling pathways are used by the salivary gland to direct its final positioning (Kolesnikov and Beckendorf, 2005). Mis-expression of *FoxL1* in all muscle activate a variety of signaling pathways in the salivary gland to

contribute to its mis-migration. During Stage 17, the glands become quite long and extend almost to the middle of the embryo. This phenotype is reminiscent of loss-of-function *biniou* mutants (Vining et al 2005). Biniou, another Fox family member, specifies the longitudinal visceral mesoderm, which migrates anteriorly to separate the salivary glands from the circular visceral mesoderm. *bin* mutants have long salivary glands because the gland maintains contact with the cVM and becomes elongated as the gut retracts. Mis-expression *FoxL1* may phenocopy *bin* mutants by disrupting the differentiation of the gut, thus driving gland elongation. In support of this, the guts of Twi>*FoxL1* embryos contain lesions and do not express markers for the second gut constriction. Markers for gut differentiation, such as Teashirt, should be tested in Twi>*FoxL1* embryos to test this hypothesis.

One target of FoxL1 that we identified is *sema2a*. *sema2a* is expressed in VIS5 of T3 (Kolodkin et al 1993). The Semaphorin family is a class of cysteine-rich signaling molecules that act as repellents for axons (Cohen et al 2005). Five semaphorins are encoded in *Drosophila*, and Sema2a is one of two secreted members of the family. Over-expression of *sema2a* repels neurons, and this phenotype is mediated through its receptor PlexinB (Ayoob et al 2006). Interestingly, loss-of-function of *sema2a* does not significantly impact neuron guidance, but loss of *plexB* does. The salivary gland expresses both *plexB* and *sema2a*, which was revealed through the $sema2a^{P0321}$ lacZ reporter. Similarly to the neurons, loss of *sema2a* does not significantly affect salivary gland migration, but knock-down of *plexB* does. This finding suggests that PlexB has ligands in addition to Sema2a, ligands whose expression is independent of FoxL1. We plan to perform a microarray using $fd64a^{KO}$ and $fd64a^{10.5}$ in trans to one another to identify these targets.

REFERENCES

Abrams, E.W., Mihoulides, W.K., and Andrew, D.J. (2006). Fork head and Sage maintain a uniform and patent salivary gland lumen through regulation of two downstream target genes, PH4αSG1 and PH4αSG2. *Development* **133**(18): 3517-3527.

Andrew D.J., Baig A., Bhanot P., Smolik S.M., and Henderson K.D. (1997). The *Drosophila* dCREB-A gene is required for dorsal/ventral patterning of the larval cuticle. Development. **124**:181–193

Ayoob, J. C., Terman, J. R., and Kolodkin, A. L. (2006). Drosophila Plexin B is a Sema-2a receptor required for axon guidance. *Development,* **133**(11), 2125–2135.

Baena-Lopez, L.A., Alexandre, C., Mitchell, A., Pasakarnis, L., and Vincent, J-P. (2013) Accelerated homolgous recombination and subsequent genome modification in *Drosophila*. *Development*, **140**, 4818-25.

Bate, M. (1990) The embryonic development of larval muscles in Drosophila. *Development*, **110**(3),791-804.

Baylies, M.K., Bate, M., Ruiz Gomez, M. (1998). Myogenesis: a view from Drosophila. *Cell* **93**:921–927.

Beckett, K., and Baylies, M.K. (2007). 3D analysis of founder cell and fusion competent myoblast arrangements outlines a new model of myoblast fusion. *Dev Biol* **309**:113–25.

Chung, S.-Y., Hanlon, C.D., and Andrew, D.J. (2014) Building and specializing epithelial tubular organs: the Drosophila salivary gland as a model system for revealing how epithelial organs are specified, form and specialize. *Wiley Interdiscip Rev Dev Biol.* **3**:281-300.

Cohen, S., Funkelstein, L., Livet, J., Rougon, G., Henderson, C.E., Castellani, V., and Mann, F. (2005). A semaphorin code defines subpopulations of spinal motor neurons during mouse development. *Eur J Neurosci* **21**(7), 1767–76.

Coulson, M., Robert, S., and Saint, R. (2005). Drosophila starvin encodes a tissue-specific BAG-domain protein required for larval food uptake. *Genetics*, **171**(4), 1799–1812.

Folker, E.S., Schulman, V.K., and Baylies, M.K. (2012). Muscle length and myonuclear position are independently regulated by distinct Dynein pathways. *Development*, **139**(20), 3827-37.

Fox, R. M., Vaishnavi, A., Maruyama, R., and Andrew, D. J. (2013). Organ-specific gene expression: the bHLH protein Sage provides tissue specificity to Drosophila FoxA. *Development*, ***140***(10), 2160–71.

Frasch M. (1999). Controls in patterning and diversification of somatic muscles during Drosophila embryogenesis. *Curr Opin Genet Dev.* **9**:522–529.

Ghazi, A., Paul, L., VijayRaghavan, K. *(*2003*).* Prepattern genes and signaling molecules regulate stripe expression to specify *Drosophila* flight muscle attachment sites*. Mech Dev* **120***:* 519–528.

Gong, W.J., and Golic, K.G. (2003). Ends-out, or replacement, gene targeting in Drosophila. *PNAS* **100**(5): 2556-2561.

Greig, S., and Akam, M. (1993). Homeotic genes autonomously specify one aspect of pattern in the Drosophila mesoderm. *Nature* **362**(6421): 630-632.

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E.E., Lints, T, Jarvis, E.D., and Scharff, C. (2004). FoxP2 expression in avian vocal learners and non-learners. *J Neuorsci* **24**(13): 3164-75.

Hannenhalli, S., and Kaestner, K. H. (2009). The evolution of Fox genes and their role in development and disease. *Nature Reviews Genetics*, **10**(4), 233–240.

Harris K.E., Schnittke N., and Beckendorf S.K. (2009). Two ligands signal through the Drosophila PDGF/VEGF Receptor to Ensure Proper Salivary Gland Positioning. *Mech Dev* **124**(6):441-448.

Henderson, K.D., and Andrew, D.J. (2000). Regulation and function of Scr, exd, and hth in the Drosophila salivary gland. *Dev. Biol.* **217**(2): 362-374.

Jackson, B. C., Carpenter, C., Nebert, D. W., and Vasiliou, V. (2010). Update of human and mouse forkhead box (FOX) gene families. *Human Genomics*, **4**(5), 345–352.

Katz, J. P., Perreault, N., Goldstein, B. G., Chao, H.-H., Ferraris, R. P., and Kaestner, K. H. (2004). Foxl1 null mice have abnormal intestinal epithelia, postnatal growth retardation, and defective intestinal glucose uptake. *American Journal of Physiology. Gastrointestinal and Liver Physiology*, **287**(4), G856–G864.

Kiehard, D.P., and Feghali, R. (1986). Cytoplasmic Myosin from *Drosophila melanogaster*. *JCB* **103**, 1517-24.

Kolesnikov, T., and Beckendorf, S.K. (2005). NETRIN and SLIT guide salivary gland migration. *Dev Biol*. **284**(1):102-111.

Kolodkin, A.L., Matthes, D.J., Goodman, C.S. (1993). The semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules. *Cell* **75**(7):1389-1399.

Kocherlakota, K.S., Wu, J.M., McDermott, J., and Abmayr, S.M. (2008). Analysis of the cell adhesion molecule sticks-and-stones reveals multiple redundant functional domains, protein-interaction motifs and phosphorylated tyrosines that direct myoblast fusion in Drosophila melanogaster. *Genetics* **178**(3): 1371-1383.

Kramer, S.G.*,* Kidd, T.*,* Simpson, J.H.*, and* Goodman, C.S. *(*2001*).* Switching repulsion to attraction: changing responses to slit during transition in mesoderm migration*. Science* 292*:* 737–740

Lee, H. H., and Frasch, M. (2004). Survey of Forkhead Domain Encoding Genes in the Drosophila Genome: Classification and Embryonic Expression Patterns. *Developmental Dynamics*, **229**(2), 357–366.

Lehmann, R., and Tautz, D. (1994). In situ hybridization to RNA. *Mehtods Cell Biol* **44**: 575-98.

Maruyama, R., Grevengoed, E., Stempniewicz, P., Andrew D.J. (2011). Genome-wide analysis reveals a major role in cell fate maintenance and an unexpected role in endoreduplication for the Drosophila FoxA gene fork head. *PLoS One* **6**(6).

Menon, S.D., and Chia, W. (2001). Drosophila rolling pebbles: a multidomain protein required for myoblast fusion that recruits D-Titin in response to the myoblast attractant Dumbfounded. *Dev. Cell* **1**(5): 691-703.

Myat, M.M., Andrew, D.J. (2000). Fork head prevents apoptosis and promotes cell shape change during formation of the Drosophila salivary glands. Development **127**(19), 4217-4226.

Myatt, S. S., and Lam, E. W.-F. (2007). The emerging roles of forkhead box (Fox) proteins in cancer. *Nature Reviews. Cancer*, **7**(11), 847–859.

Perreault, N., Katz, J. P., Sackett, S. D., and Kaestner, K. H. (2001). Foxl1 Controls the Wnt/??-Catenin Pathway by Modulating the Expression of Proteoglycans in the Gut. *Journal of Biological Chemistry*, **276**(46), 43328–43333.

Reuter, R., and Scott, M.P. (1990). Expression and function of the homoeotic genes Antennapedia and Sex combs reduced in the embryonic midgut of Drosophila. *Development* **109**:289–303.

Richardson, B., Beckett, K., and Baylies, M. (2008). Visulaizing new dimensions in Drosophila myoblast fusion. *Bioessays* **30**, 423-431.

San Martin, B., and Bate M. (2001). Hindgut visceral mesoderm requires an ectodermal template for normal development in Drosophila. *Development* **128**(2), 233-242.

Tremml, G., and Bienz, M. (1992). Induction of labial expression in the Drosophila endoderm: response elements for dpp signalling and for autoregulation. **Development** *116*(2): 447-456.

Vining, M.S., Bradley, P.L., Comeaux, C.A., and Andrew, D.J. (2005). Organ positioning in Drosophila requires complex tissue-tissue interactions. *Dev Biol*. **287**(1):19-34.

Weigel, D.*,and* Jackle, H. *(*1990). The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? Cell **63***:* 455–456*.*

Weigel, D., Jurgens, G., Kuttner, F., Seifert, E.*, and* Jackle H. *(*1989). The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo*.* Cell **57***:* 645–658*.*

Wu, Z., Sweeney, L. B., Ayoob, J. C., Chak, K., Andreone, B. J., Ohyama, T., Kerr, R., Luo, L., Zlatic, M., and Kolodkin, A. L. (2011). A Combinatorial Semaphorin Code Instructs the Initial Steps of Sensory Circuit Assembly in the Drosophila CNS. *Neuron*, **70**(2), 281–298.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes and Development* **25**(21): 2227-41.

Chapter 6:

The CrebA/Creb3-like transcription factors are major and direct regulators of secretory capacity

This chapter is modified from: Fox, R.M., Hanlon, C.D., and Andrew, D.J. (2010). The CrebA/Creb3-like transcription factors are major and direct regulators of secretory capacity. *Journal of Cell Biology* 191: 479-492.

Abstract

Secretion occurs in all cells, with relatively low levels in most cells and extremely high levels in specialized secretory cells, such as those of the pancreas, salivary and mammary glands. How secretory capacity is selectively upregulated in specialized secretory cells is unknown. Here, we report that the CrebA/Creb3-like family of bZip transcription factors functions to upregulate expression of both the general protein machinery required in all cells for secretion and of cell-type specific secreted proteins. Drosophila CrebA directly binds the enhancers of secretory pathway genes and is both necessary and sufficient to activate expression of every secretory pathway component gene examined thus far. Microarray profiling reveals that CrebA also upregulates expression of genes encoding cell type-specific secreted components. Finally, we find that the human CrebA orthologues, Creb3L1 and Creb3L2, have the ability to upregulate the secretory pathway in non-secretory cell types.

INTRODUCTION

The human pancreas secretes liters of enzymes daily to aid in food digestion, whereas the bovine mammary glands produce eight liters of milk each day, largely for human consumption. To do this, secretory organs must adapt to the increased need for protein secretion that occurs during development, differentiation, or changing physiological conditions. An important question is how changes in secretory capacity are coordinated to allow for efficient targeting, folding, modification and delivery of secreted products. A few transcription factors have been discovered to upregulate genes in the secretory pathway including Xbp1, which is expressed and required in B cells as they differentiate into antibody secreting plasma cells (Shaffer et al., 2002) and which also regulates secretory function in a subset of specialized secretory organs (Lee et al., 2005; Shaffer et al., 2004). The bZip transcription factor ATF6 activates expression of chaperone proteins required for efficient protein folding (Adachi et al., 2008) as well as many of the lipid components of secretory organelles (Bommiasamy et al., 2009). Two other bZip transcription factors, Creb3L1/OASIS and Creb3L2/BBF2H7 (herein referred to as Creb3L1 and Creb3L2), are required for efficient bone deposition and cartilage matrix secretion, respectively (Murakami et al., 2009; Saito et al., 2009). A major question is whether these transcription factors function more broadly to upregulate the entire secretory pathway in multiple specialized cell types or if their function is restricted to the upregulation of only a subset of secretory genes in a few specialized cells.

The Drosophila salivary gland (SG) provides an excellent model for identifying and studying the factors required for secretory function. The SG is the largest secretory organ in Drosophila and the processes of morphogenesis and differentiation have been well characterized (Kerman et al., 2006). The SG comprises two large secretory tubes, each containing ~100 polarized epithelial cells that are specialized for the production and delivery of secreted proteins.

Consistent with the high level secretory activity of the SG, at least 34 secretory pathway component genes (SPCGs) are highly expressed in the secretory cells (Abrams and Andrew, 2005), and this expression requires at least two transcription factor genes, *fork head* (*fkh*) and *CrebA* (Andrew et al., 1997; Myat et al., 2000).

Salivary gland expression of *fkh* and *CrebA* is activated in the most posterior head segment, (parasegment two), by the homeotic gene *Sex combs reduced* (*Scr*) and two more generally expressed homeotic cofactor genes *extradenticle* (*exd*) and *homothorax* (*hth*) (Henderson and Andrew, 2000). Dpp signaling in dorsal cells blocks expression of *fkh* and *CrebA*, limiting their activation to only the ventral cells of parasegment two (Henderson et al., 1999). Shortly after activation of *fkh* and *CrebA*, expression of *Scr, exd* and *hth* disappears in the salivary gland (Henderson and Andrew, 2000); continued expression of both *fkh* and *CrebA* is maintained by Fkh (Abrams and Andrew, 2005). Thus, we propose that Fkh plays a primarily indirect role in SPCG expression through its role in maintaining expression of *CrebA* (Abrams and Andrew, 2005). Consistent with this idea, the loss of *fkh* affects only late SPCG expression, whereas loss of *CrebA* affects both early and late SPCG expression. It is unknown, however, if CrebA directly regulates SPCG expression or if additional downstream factors are also involved.

Here, we show that CrebA is both necessary and sufficient for high level SPCG expression in the secretory tissues of the Drosophila embryo. We show that direct binding of CrebA to a consensus motif identified upstream of the 34 originally characterized SPCGs is required for elevated SPCG expression in the secretory tissues. Through microarray analysis, we find that over half of the 383 genes that require CrebA encode identifiable secretory pathway components. Surprisingly, CrebA targets include not only components of the general secretory machinery that function in all cells but also cell-type specific secreted cargo. Moreover, phenotypes associated with loss of *CrebA* are consistent with the role of this gene in secretion.

Finally, we confirm Creb3L1 and Creb3L2 as the closest mammalian orthologues to Drosophila CrebA and demonstrate that both human proteins have the same activities as their Drosophila counterpart.

MATERIALS AND METHODS

Fly Strains

The *CrebA*^{wR23} protein null allele was used for all loss-of-function analysis (Andrew et al., 1997). UAS-CrebB and *CrebB*^{S162} were obtained from the Bloomington Stock Center: HS-FLP and P{ovo^{D}} FRT2A (79D)/TM3 were obtained from Elizabeth Chen. Recessive lethal lines were balanced over *lacZ* or GFP balancers, to allow identification of homozygous mutants. UAS-Creb3L1-FL/T and UAS-Creb3L2-FL/T were generated by Gateway (Invitrogen, Carlsbad, CA) mediated cloning and recombination into the pTW untagged UAS vector (Drosophila Gateway Collection, Carnegie Institute, Baltimore, MD). *engrailed (en)*-Gal4, and *breathless (btl)*-Gal4 were used to express UAS-CrebA (Rose et al., 1997), UAS-Creb3L1 FL, UAS-Creb3L1 T, UAS-Creb3L2 FL or UAS-Creb3L2 T constructs in epidermal stripes (*en*-Gal4) (Weiss et al., 2001) or the trachea, salivary duct and midline (*btl*-Gal4) (Shiga, 1996).

Electrophoretic mobility shift assays

Plus and minus strand oligonucleotides of ~30 bases were designed and synthesized (IDT DNA; Coralville, IA) for each binding site and included the nine nucleotide consensus motif flanked by ~10 nucleotides of genomic sequence at both the 5' and 3' ends. The plus strand of each oligo was labeled with □^{32}P using the manufacturer's protocol (Invitrogen, Carlsbad, CA) and was annealed to the minus strand by heating the DNA to 95°C for 5 min. and then cooling to room temperature (RT). Unlabeled double stranded oligos were prepared similarly for competition experiments. DNA binding reactions were performed as described (Smolik et al., 1992) with the exception that all competitor oligonucleotides were added at concentrations of 20, 60 and 100X. Binding reactions were run on a 4% polyacrylamide gel at 30 mAmps for ~2 hours and prepared for autoradiography using standard methods.

Site-directed mutagenesis of SPCG enhancer-lacZ reporters

Primers for putative CrebA binding sites (Abrams and Andrew, 2005) were designed to mutate the five core nucleotides of the CrebA binding motif G/TACGT to ACAAC using the Stratagene Primer Design program. Sites were subsequently mutated using the QuikChange Multi Site-Directed Mutagenesis kit (Stratagene, Cedar Creek, TX) and the mutated constructs were injected into w^{1118} flies (Rainbow Transgenics, Newbury Park, CA).

Chromatin-Immunoprecipitation and quantitative PCR

The Chromatin IP followed the protocol of Birch-Machin et al. ((Birch-Machin et al., 2005). Quantitative real time PCR was performed using 1 □l of eluted sample DNA, 1.5 □M forward and reverse primers, and iQ SYBR Green Supermix (BioRad). Experiments were performed in triplicate, including those from an input control. Samples were normalized to *actin5c* and fold change over the no-primary antibody control was calculated using the ddCt method (Livak and Schmittgen, 2001). Error bars represent the standard deviation of ddCt.

Transmission electron microscopy

CrebA mutant embryos were identified by the absence of GFP staining from the TM3 balancer chromosomes. Stage 16 wild-type and *CrebA* mutant embryos were processed for TEM using standard protocols and examined on a Philips EM120 microscope. To determine secretory vesicle size, area measurements were collected using ImageJ software (NIH), and statistical significance was determined using a two-tailed Student's t-test.

Immunohistochemistry and in situ hybridizations

In situ hybridization and immunohistochemistry were performed as previously described (Lehmann and Tautz, 1994; Reuter et al., 1990). Antibody concentrations used in this study are as follows: α-CrebA (1:1,000) (Andrew et al., 1997), α-Crumbs (1:100) (Developmental Studies Hybridoma Bank, Iowa City, IA) and αβgalactosidase (Promega, 1:10,000). All secondary antibodies (Vector Labs and Molecular Probes) were used at a 1:500 dilution. The fluorochrome Alexa 568 was used for HeLa cell experiments, and nuclei were labeled with DAPI. Confocal images were obtained using a Zeiss LSM 510 Meta confocal microscope, using a Plan-Neofluor 40x, 1.3 oil objective and the Zeiss LSM software. All other images were obtained using a Zeiss Axiophot microscope configured with a Nikon Coolpix 4500 digital camera. Images were taken using a Plan-Neofluor 20x, 0.50 objective. Images were rotated and cropped using Adobe photoshop. All images were obtained at room temperature.

Generation of CrebA maternal-zygotic mutants

CrebA^{23w-} was recombined onto a third chromosome containing FRT2A (79D) and then crossed to flies carrying hsFLP; P{ovo^{D}} FRT2A (79D). After 72 hours, larvae were heat-shocked for one hour at 37°C on consecutive days until pupae formation. Female flies carrying both *CrebA*$^{23w-}$ FRT2A(79D) and P{ovo^{D}} FRT2A (79D) were crossed to *CrebA*$^{23w-}$*/TM6B, Ubx-lacZ* males, and progeny were collected and stained using the appropriate markers. *CrebA*$^{mat-zyg-}$ embryos were identified based on the absence of staining for gal from the lacZ-containing balancer.

Microarray experiments to identify CrebA target genes in Drosophila

Three samples of sorted stage 11-16 *CrebA* homozygous (GFP balancer negative) mutant embryos and wild-type OR embryos were isolated using a COPAS Select embryo sorter (Union Biometrica). Total RNA was isolated using TriZOL:Chloroform extraction and precipitated with isopropanol. The Qiagen RNeasy kit was used for RNA clean up. Total RNA (100ng) was labeled according to standard Affymetrix protocols and hybridized to the Drosophila genome 2.0 chip. Following scanning, intensity values were normalized by RMA (Irizarry et al., 2003a; Irizarry et al., 2003b) (Partek Inc., St. Louis, MO) and statistical analysis was performed using Spotfire software (TIBCO, Palo Alto, CA). CrebA target genes were identified based on a 1.5 fold change in gene expression, with a p-value $\leq$ 0.05. This fold change was selected since many known, confirmed, CrebA target genes fall into the range from 1.5-2.0x.

HeLa cell culture, transfection, and immunofluorescence

cDNAs corresponding to the full length Creb3L1 and Creb3L2 were purchased from Open Biosystems (Accession numbers BC01578 and BC110813, respectively) and the full length and truncated (Creb3L1: amino acids 1-375, Creb3L2: amino acids 1-379) versions were cloned into pCDNA3.1/Hygro (Invitrogen) for transient transfection. DNA (1 µg of Creb3Lx with 0.1 µg of the co-transfection marker pEGFP-C1 [Invitrogen]) was transfected into cells using the Fugene6 transfection reagent according to the manufacturer's protocol (Roche). Immunofluorescent staining was performed using established protocols (Sbodio and Machamer, 2007). Primary antibodies for Creb3L1 (Accession number NP_443086) and Creb3L2 (Accession number NP_919047) were purchased from Aviva Systems Biology (San Diego, CA), and used at a dilution of 1:500. Fluorescent secondary antibodies (Molecular Probes) were used at a dilution of 1:500. Co-staining with an ER marker (DM286, gift from Chris Nicchitta) and the nuclear marker DAPI,

confirmed both the ER localization and nuclear localization of the full length and truncated Creb3L proteins, respectively (data not shown).

Cell sorting, RNA extraction and microarray analysis from HeLa cells

HeLa cells were co-transfected with both Creb3L1 T and EGFP approximately 20 hours prior to sorting (as above). As a control, HeLa cells were mock transfected with Fugene6 reagent, and sorted using identical conditions. Cells were sorted on a FACSaria flow cytometer, and at least 200,000 cells were isolated. RNA was extracted using the Qiagen RNeasy kit. Total RNA (100ng) was labeled and hybridized to the Affymetrix Human Gene 1.0 array according to standard Affymetrix protocols. Following scanning, raw data was normalized by RMA (Irizarry et al., 2003a; Irizarry et al., 2003b) (Partek Inc., St. Louis, MO) and statistical analysis was performed using Spotfire software (TIBCO, Palo Alto, CA).

Accession numbers

All microarray data has been deposited in Gene Expression Omnibus (GEO), accession number GSE23349.

RESULTS

CrebA binds directly to SPCG enhancers in vitro and in vivo

Drosophila CrebA expression is elevated in many secretory organs in the embryo, with highest expression in the developing SG, proventriculus, late trachea and epidermis (Fig. 1A; Andrew et al., 1997). In these tissues, CrebA is required for the high level expression of 34 known secretory pathway component genes (SPCGs) (Abrams and Andrew, 2005). A MEME analysis (http://meme.sdsc.edu/meme/website) of the enhancer regions upstream of these genes revealed a conserved motif similar to the previously characterized CREB response element (CRE) (Montminy and Bilezikjian, 1987) and unfolded protein response elements (UPRE) (Wang et al., 2000) that bind the mammalian CREB proteins (Fig. 2A) (Abrams and Andrew, 2005). To ask if the more distantly related CrebA protein binds these sites *in vitro*, we performed electrophoretic mobility shift assays (EMSA) with ~30 nucleotide double-stranded oligomers corresponding to 18 putative CrebA binding sites found within the CrebA-dependent enhancers of five representative SPCGs: *SrpRα*, *Sec61β*, *Spase25*, *p24.1*, and *δCop* (Fig. 1C). Each gene encodes a protein found in a distinct complex functioning at a different step in early secretion (Fig. 1B). Using purified CrebA protein, we observed strong binding in all cases, as revealed by the decreased mobility of CrebA:DNA complexes relative to unbound DNA (Fig. 2B; Fig. S1). The binding is specific; whereas unlabeled competitor oligomers corresponding to the same sequence as the labeled probe competed for CrebA binding in 17 of the 18 sites tested, unlabelled competitor oligos in which the five core nucleotides of the consensus motif were changed did not compete (Fig. 2B, Fig. S1). We then performed Chromatin Immunoprecipitation (ChIP) followed by quantitative RT-PCR to determine if CrebA binds the enhancer regions upstream of these SPCGs *in vivo*. Chromatin was extracted from 0-24 hr embryos and immunoprecipitated with either CrebA antiserum or CrebA preimmune serum as a negative

Figure 1. CrebA is expressed in secretory tissues and regulates SPCG expression. (A). Endogenous CrebA protein is detected in the salivary glands (white asterisks), epidermis (arrowheads), and proventriculus (black arrow) beginning at stage 11 (lateral view) and continuing through stage 16 of embryonic development (images of stage 13, 15 and 16 are ventral views). Elevated expression is also detected in the trachea, but is not visible in these images. (B) Cartoon representation of the protein complexes represented by the five enhancers characterized in this study. (C) *In situ* hybridizations show loss of SG (black asterisks), proventriculus (white arrow) and epidermal expression (arrowheads) in the *CrebA* mutant. All scale bars = 125μm.

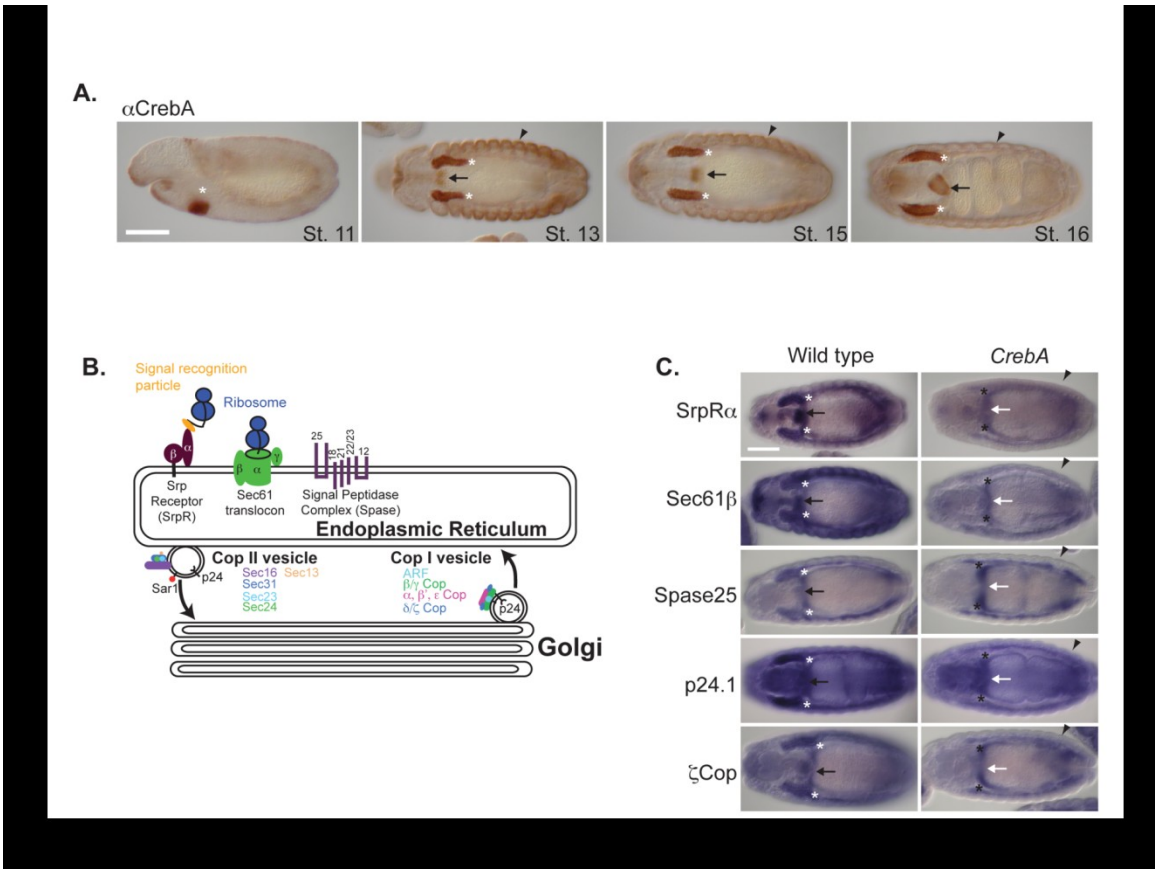
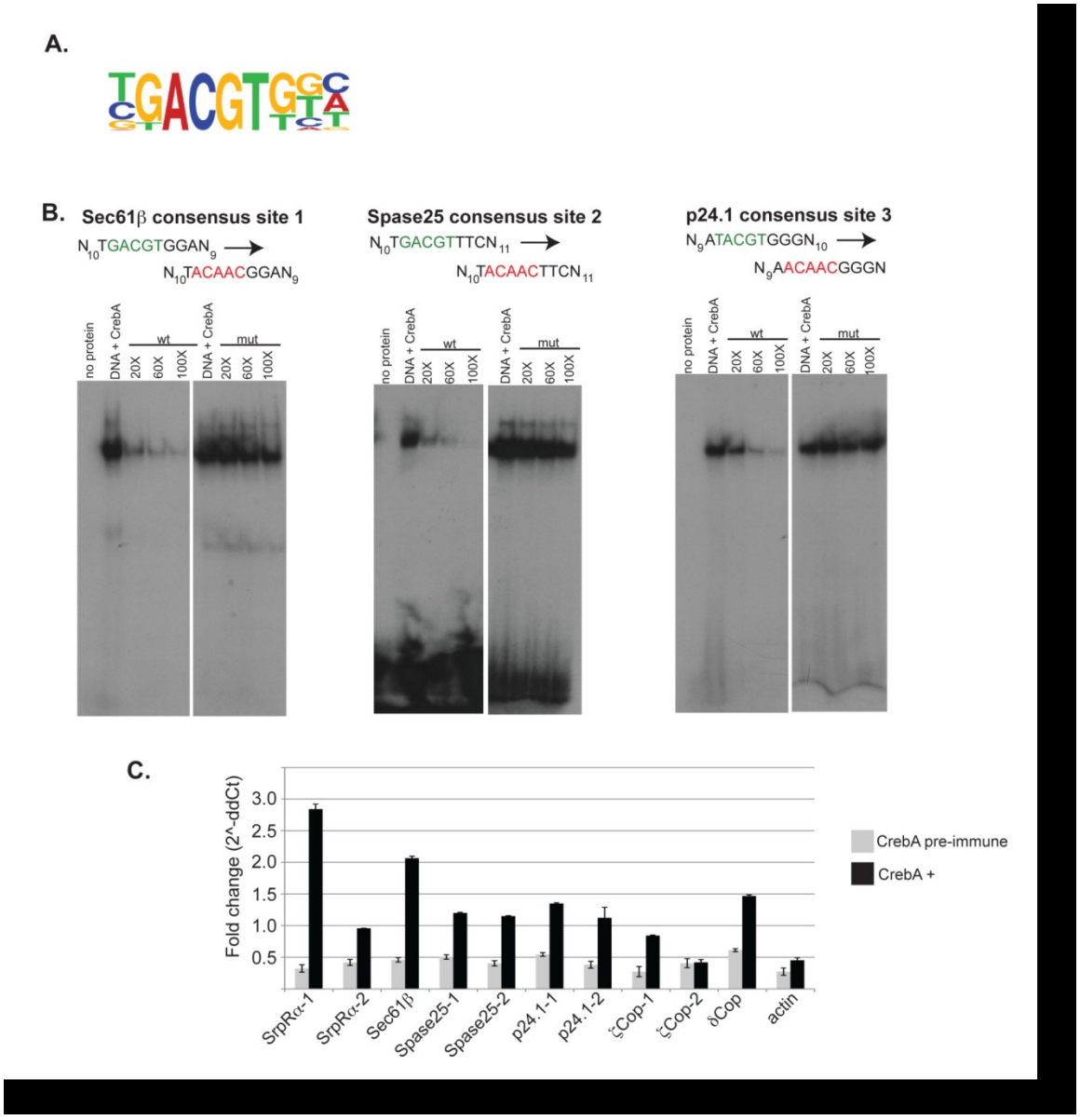

A.
αCrebA
St. 11
St. 13
St. 15
St. 16
B.
Signal recognition particle
Ribosome
Srp Receptor (SrpR)
Sec61 translocon
Signal Peptidase Complex (Spase)
25
22/23
12
Endoplasmic Reticulum
Cop II vesicle
Sar1
p24
Sec16
Sec13
Sec31
Sec23
Sec24
Cop I vesicle
ARF
βγ Cop
α, β', ε Cop
δ/ζ Cop
p24
Golgi
C.
Wild type
CrebA
SrpRα
Sec61β
Spase25
p24.1
ζCop

Figure 2. CrebA directly activates SPCG expression. (A) The CrebA consensus motif identified upstream of 34 SPCGs by MEME analysis (Abrams and Andrew, 2005). (B) EMSAs reveal that CrebA binds to radio-labeled double stranded (ds) oligonucleotides containing the CrebA consensus motif (green sequence). Unlabeled wild type (wt) ds oligonucleotides compete for CrebA binding, whereas unlabeled ds oligonucleotides containing mutated consensus motifs (red sequence, mut), do not. (C) Chromatin-Immunoprecipitation (ChIP) analysis using CrebA specific antibodies, followed by quantitative PCR, reveals that SPCG enhancer regions exhibit increased CrebA occupancy when compared to control genes. This figure shows the average from one representative trial (out of three). Of note, Spase25-2 showed lower occupancy in one trial, while Cop-2 showed higher occupancy in the other two trials. ddCt indicates the relative abundance of a gene (or DNA fragment) normalized to a housekeeping gene (*actin5c*) and normalized to an IgG control. Error bars represent standard error of the mean.


A.
B.
Sec61β consensus site 1
N10TGACGTGGAN9
N10TACAACGGAN9
Spase25 consensus site 2
N10TGACGTTTCN11
N10TACAACTTCN11
p24.1 consensus site 3
N9ATACGTGGGN10
N9AACAACGGGN
no protein
DNA + CrebA
wt
mut
20X
60X
100X
C.
Fold change (2^-ddCt)
3.0
2.5
2.0
1.5
1.0
0.5
SrpRα-1
SrpRα-2
Sec61β
Spase25-1
Spase25-2
p24.1-1
p24.1-2
ζCop-1
ζCop-2
δCop
actin
CrebA pre-immune
CrebA +

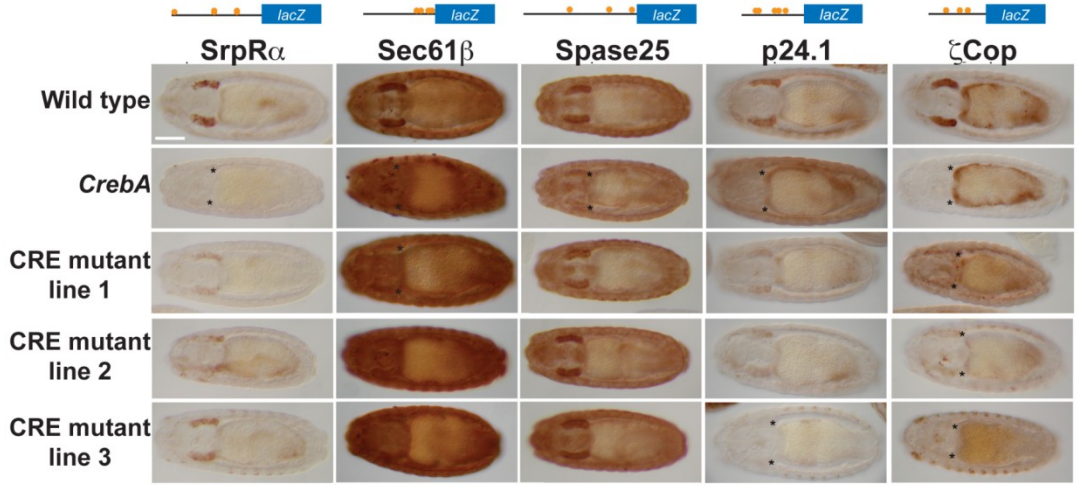
control. All tested SPCG enhancers were preferentially pulled down with the CrebA antibody (Fig. 2C). These findings indicate that CrebA binds to the conserved CrebA consensus site both *in vitro* and *in vivo*.

CrebA directly activates SPCG expression

To ask if the CrebA binding sites are required for CrebA-dependent expression of SPCGs *in vivo*, we generated multiple independent transgenic lines in which all of the CrebA consensus motifs within the SPCG enhancers were mutated. Transgenic lines carrying wild-type SPCG enhancers driving *lacZ* had robust □gal expression in the SGs of wild-type embryos, which was significantly reduced in *CrebA* mutants (Abrams and Andrew, 2005). When stained in parallel with the wild-type constructs, most lines carrying mutated CrebA binding sites for *SrpRα*, *p24.1*, *Sec61β*, and *δCop* had significantly reduced levels of SG βgal expression when visualized in otherwise wild-type embryos (Fig. 3). Surprisingly, most of the mutant *Spase25* enhancer lines (8/10) had levels of □gal expression similar to the wild-type *Spase25* enhancer. The residual SG □gal expression observed with the mutated enhancers of *SrpRα*, *p24.1*, *Sec61β*, and *δCop* and the strong βgal expression from the mutated enhancer of *Spase25* suggests that although CrebA directly activates expression of most SPCGs through the sites that bind CrebA *in vitro*, there may also be some activation by CrebA-dependent downstream transcription factors or CrebA may also directly activate SPCG expression through divergent binding sites still contained within the enhancer regions of the mutated SPCGs. Altogether, these results indicate that CrebA is required for SPCG expression, and that full expression of most SPCGs is dependent on the CrebA consensus binding motif identified through the MEME analysis.

Figure 3. SPCG expression *in vivo* requires the CrebA consensus motif. (A) The structures of the lacZ reporter gene constructs are represented above each gene name with putative CrebA binding sites represented by orange circles. LacZ reporter lines show high level SG expression in wild-type embryos (top panels), which is significantly reduced in *CrebA* mutants (second row panels). Mutation of the CrebA response element (CRE) results in significant reduction of βGal expression for *SrpRα*, *Sec61β*, *p24.1*, and *δCop* but not for *Spase25* (lower panels) in otherwise WT embryos. Three individual transgenic lines are shown for each reporter. Asterisks denote SG location in the mutant lines. Scale bars = 125μm. (B) Table showing the number of independent transgenic lines analyzed for each mutated enhancer and the relative level of lacZ expression compared to the wild-type reporter lines.

A.



B.

Enhancer	*Number of lines*	*Reduced*	*Elevated*	*No Change*
SrpRα	7	6	1	0
Sec61β	11	11	0	0
Spase25	10	2	0	8
p24.1	8	5	1	2
ζCop	12	10	0	2

CrebA is sufficient to induce SPCG expression

We next asked if CrebA can induce high level SPCG transcription in cells that normally only express low or undetectable levels of these genes. UAS-CrebA (Rose et al., 1997) was expressed using either the *breathless*-Gal4 (trachea, midline, salivary duct) or *engrailed*-Gal-4 (epidermal stripes) drivers (Shiga, 1996; Weiss et al., 2001). In wild-type embryos, *CrebA* expression is not detected in the midline cells, the salivary duct, or in the engrailed domains, and only moderate levels are detected in the trachea (Fig. 4A). Ectopic expression of CrebA resulted in elevated levels of expression of all five of the SPCGs in all locations (Fig. 4A). Neither loss nor ectopic expression of *CrebB*, the most closely related Drosophila gene, had any effect on SPCG expression (Fig. 4B and data not shown). Thus, expression of CrebA alone is sufficient to upregulate SPCG expression in multiple distinct embryonic cell types, suggesting a major role for this transcription factor in upregulating secretory capacity.

CrebA regulates additional secretory pathway genes as well as secreted cargo

To ask if CrebA is a general regulator of SPCGs, we performed microarray analyses comparing RNA from wild-type and *CrebA* mutant embryos (Fig. 5, Table S1). Not only did we observe a significant reduction in the expression levels of the majority of the SPCGs previously analyzed in *CrebA* mutant embryos (changes of 1.5 fold or greater) but we discovered that many additional genes encoding proteins known or suspected to function in secretion were similarly affected. Gene Ontology (GO) clustering was performed using DAVID, a program that weighs the enrichment of a specific GO term in a given data set relative to the frequency of that term in the entire genome (Dennis et al., 2003; Huang da et al., 2009). This analysis revealed that in the absence of CrebA, the GO terms associated with secretory pathway function (i.e. co-translational protein targeting to the membrane, secretory pathway and protein transport) are

Figure 4. CrebA is sufficient to induce SPCG expression. (A) UAS-CrebA expression in ectopic domains using either the *btl*-Gal4 or the *en*-Gal4 driver results in activation of all tested SPCGs. *btl*-Gal4 drives UAS-*CrebA* expression and, consequently, SPCG expression in the trachea and CNS midline (arrowheads) of late stage 11 embryos (ventral views, Rows 1, 2). *en*-Gal4 drives UAS-CrebA expression and, consequently, SPCG expression in epidermal stripes of early stage 11 embryos (lateral views, Rows 3, 4). (B) Ectopic expression of UAS-CrebB by the *en*-Gal4 driver does not result in SPCG expression in the *engrailed* domain. All scale bars = 125µm.

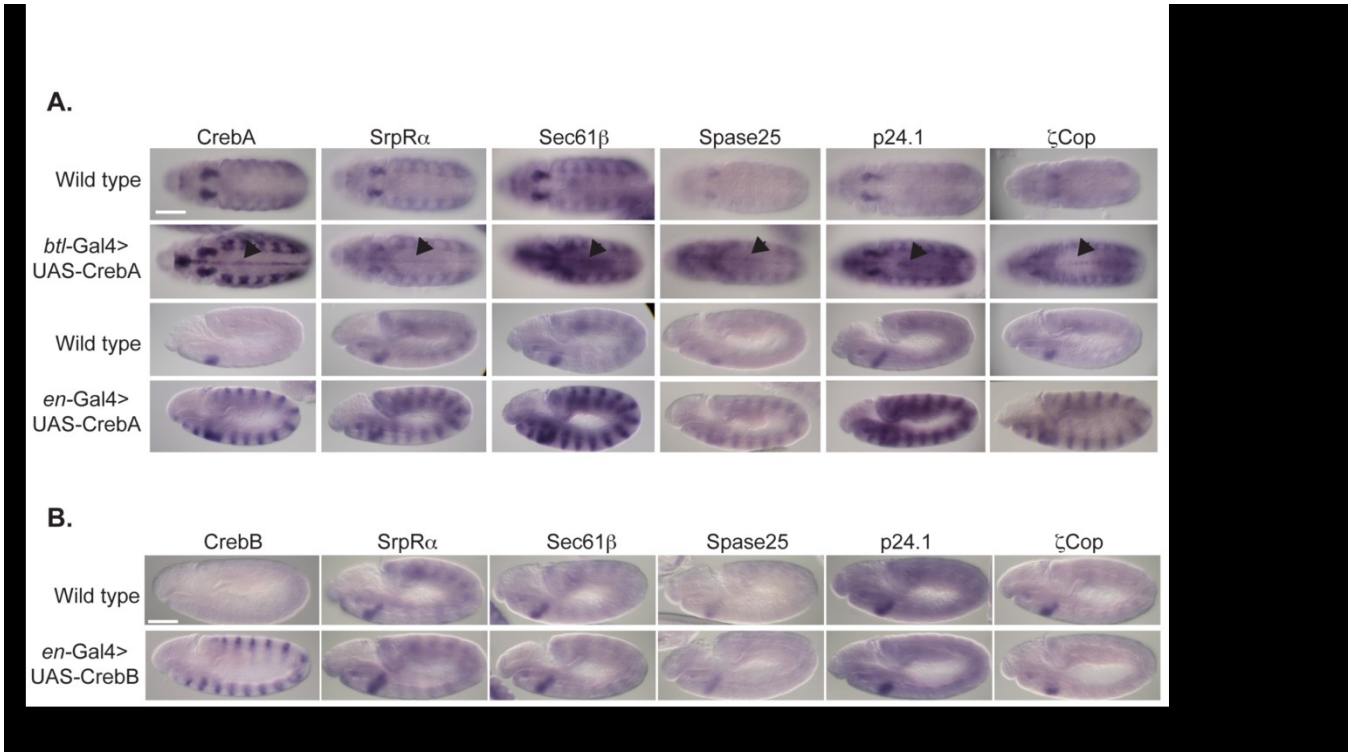
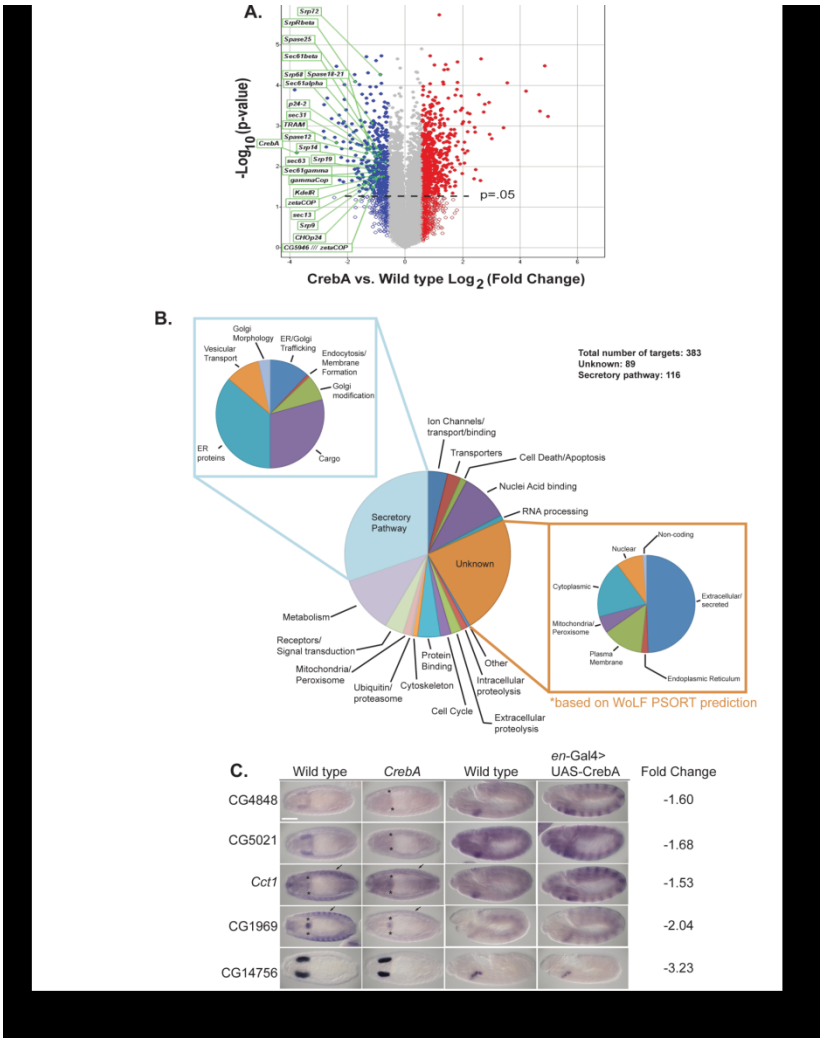

A.
CrebA
SrpRα
Sec61β
Spase25
p24.1
ζCop
Wild type
btl-Gal4>
UAS-CrebA
Wild type
en-Gal4>
UAS-CrebA
B.
CrebB
SrpRα
Sec61β
Spase25
p24.1
ζCop
Wild type
en-Gal4>
UAS-CrebB

Figure 5. CrebA activates additional secretory pathway genes as well as secreted cargo. (A) Volcano plot showing changes in expression level and statistical significance of CrebA target genes. Transcripts elevated 1.5x or more in *CrebA* mutants are labeled red. Transcripts reduced 1.5x or more are labeled blue or green. Closed circles indicate significance with a p-value $\leq$0.05. Open circles indicate non-statistically significant changes. (B) Pie charts showing predicted functions of all 383 genes downregulated in *CrebA* mutants in the microarray experiments. The 116 secretory pathway genes are further categorized according to known function and/or localization (blue box). Unknown genes (89) were further subdivided according to predicted localization as ascertained by WoLF PSORT analysis. (C) *In situ* hybridizations validate CrebA regulation of new target genes identified by microarray analysis. Overexpression of CrebA using the *en*-Gal4 driver is sufficient to induce ectopic expression of four of the five newly identified target genes. Scale bars = 125µm.


A.
-Log10 (p-value)
p=.05
CrebA vs. Wild type Log2 (Fold Change)
B.
Total number of targets: 383
Unknown: 89
Secretory pathway: 116
Secretory Pathway
Unknown
Golgi Morphology
ER/Golgi Trafficking
Vesicular Transport
Endocytosis/ Membrane Formation
Golgi modification
ER proteins
Cargo
Ion Channels/ transport/binding
Transporters
Cell Death/Apoptosis
Nuclei Acid binding
RNA processing
Metabolism
Receptors/ Signal transduction
Mitochondria/ Peroxisome
Ubiquitin/ proteasome
Cytoskeleton
Protein Binding
Cell Cycle
Other
Intracellular proteolysis
Extracellular proteolysis
Non-coding
Nuclear
Cytoplasmic
Extracellular/ secreted
Mitochondria/ Peroxisome
Plasma Membrane
Endoplasmic Reticulum
*based on WoLF PSORT prediction
C.
Wild type
CrebA
Wild type
en-Gal4> UAS-CrebA
Fold Change
CG4848
-1.60
CG5021
-1.68
Cct1
-1.53
CG1969
-2.04
CG14756
-3.23

the most highly enriched in the data set with enrichment of at least 3-fold compared to the entire Drosophila genome (Fig. 5B; Table S2). Unexpectedly, many genes predicted to encode secreted cargo also showed significantly reduced expression in *CrebA* mutants, including larval cuticle proteins, mucins and several secreted enzymes. Indeed, more than half of the ~40 genes most affected by loss of *CrebA* encode known or predicted secreted proteins, including mucins and constituents of the larval cuticle. These data indicate that CrebA not only upregulates genes encoding the general secretory machinery found in all cells but also activates genes encoding cell-type specific secreted proteins. DAVID analysis of the genes upregulated in *CrebA* mutants did not reveal enrichment for any specific pathway or biological function, suggesting that CrebA functions primarily as a transcriptional activator of genes in the secretory pathway.

In addition to known secretory pathway genes, we identified 89 CrebA target genes encoding proteins of unknown function, i.e. with no associated GO terms. A WoLF PSORT analysis (http://wolfpsort.org/) of these proteins revealed that almost half (44) contain a signal sequence and are predicted to be secreted, with an additional 14 containing transmembrane domains and predicted to localize to the ER or plasma membrane (Fig. 5B). Thus, it is likely that many of these uncharacterized genes encode either novel secretory pathway components or secreted gene products.

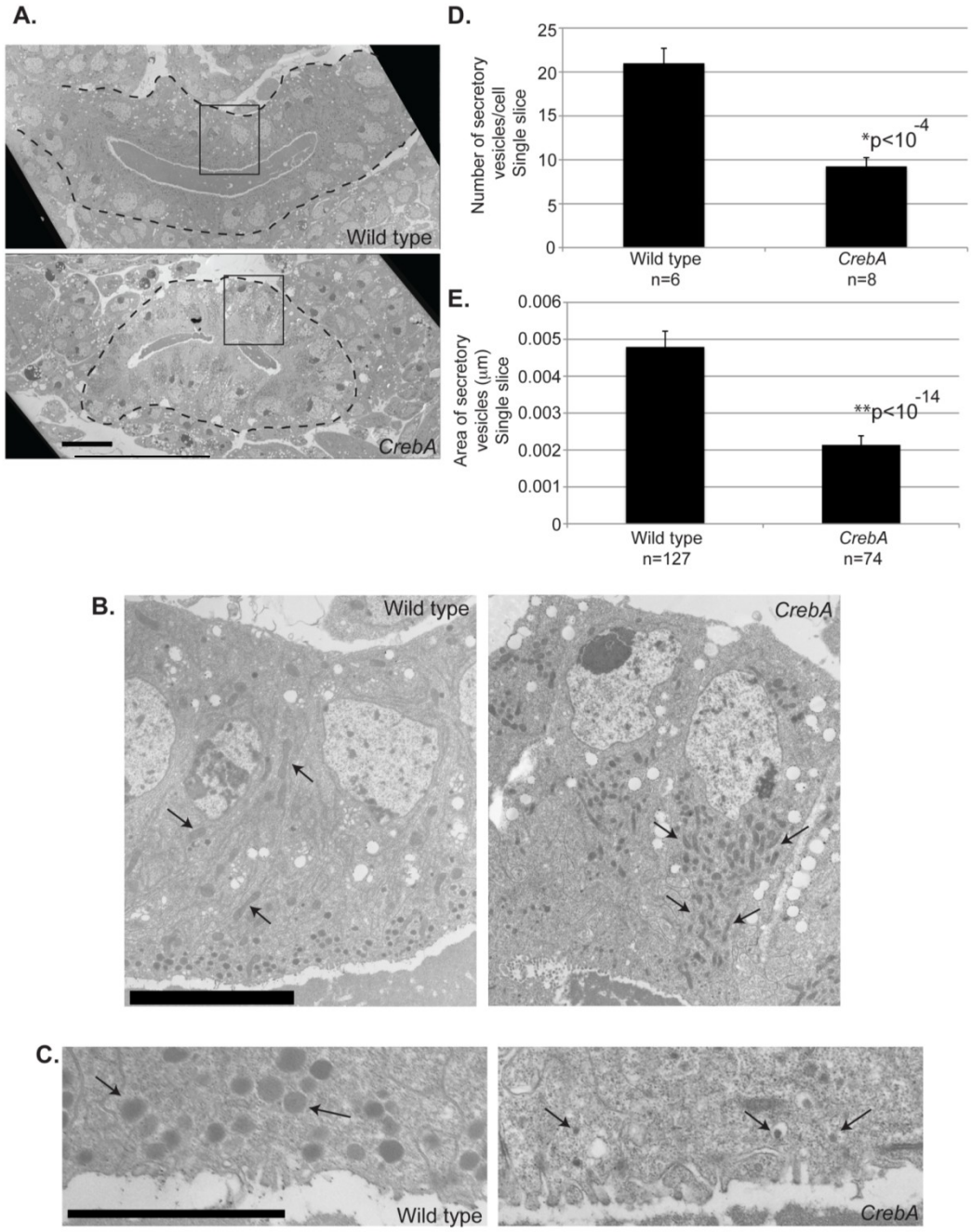
Using *in situ* hybridization, we confirmed that several genes identified as potential CrebA targets in the microarray analysis require CrebA for expression in the SG, proventriculus, and/or epidermis (Fig. 5C). Interestingly, many of them have predicted roles in the secretory pathway. Although *CG4848* does not contain GO annotation, it encodes a Vps51/Vps67 domain protein closely related to mammalian Cog1, a member of the conserved oligomeric Golgi complex required for protein glycosylation. Cct1 encodes choline phosphate cytidyltransferase, an enzyme that functions in the lipid biogenesis pathway. *CG1969* encodes a glucosamine 6-

phosphate N-acetyltransferase, a cytoplasmic enzyme involved in the metabolism of glutamate. *CG5021* and *CG14576* encode unknown proteins, but CG5021 has significant hydrophobic stretches suggesting multiple membrane spans, and CG14576 has an N-terminal signal sequence indicating that it likely travels through the secretory pathway and is secreted. Similar to the previously tested SPCGs, expression of CrebA using the *en*-Gal4 driver was sufficient to induce ectopic expression of a majority of these genes; the exception is *CG14576*, which we know requires additional cell-type specific transcription factors for its expression (see discussion). Thus, the microarray screen has revealed many new CrebA target genes known or likely to encode components of the secretory pathway as well as the specific protein products that are processed and delivered via this pathway.

Loss of CrebA leads to defects consistent with secretory dysfunction

Despite the critical role of CrebA in upregulating secretory function, CrebA mutants do not display major morphological SG defects; the SGs form normally and are only mildly crooked at late stages (Andrew et al., 1997). Transmission electron microscopy (TEM) of *CrebA* mutant SGs revealed three overt changes: (1) the lumen size was significantly smaller than that of wild type glands and was accompanied by reduced amounts of electron dense luminal material (Fig. 6A); (2) the mitochondria were concentrated to a region apical to the nucleus in contrast to wild-type SGs where mitochondria were distributed throughout the cell (Fig. 6B); (3) the secretory vesicles were much fewer and smaller than those of wild-type SGs (Fig. 6C-E). The TEM analysis also suggested reduced levels of ER, a change that could not be quantified from the TEMs. Although most of the SG defects revealed by TEM analysis are fully consistent with a role for CrebA as a general activator of secretory function, the change in mitochondrial localization was

Figure 6. Characterization of CrebA mutant SGs reveals decreases in secretion and changes in organelle positioning. (A) Low magnification (980x) TEM images of stage 15 SGs (wild-type, top; *CrebA* mutant, bottom) show a reduction in lumen size and content in the *CrebA* mutant (left panels; scale bar = 10□m). (B) Higher magnification images (2850x) of individual cells, with arrows indicating the position of the mitochondria (right panels; scale bar = 2.5□m). Note the clustering of mitochondria in the bottom panel (*CrebA* mutant). (C) High magnification (15000x) images of the apical regions of a single SG cell with arrows pointing to secretory vesicles (wild-type, top; *CrebA* mutant, bottom; scale bar = 1□m). (D) Quantification of secretory vesicle number per cell in an individual slice for wild type and *CrebA* mutant SGs (top). Average area of the secretory vesicles, determined using ImageJ software, for an individual slice, for WT and *CrebA* mutant SG cells (bottom). P-values in D were determined using a two-tailed Student's t-test.


A.
Wild type
CrebA
D.
Number of secretory vesicles/cell Single slice
25
20
15
10
5
0
*p<10-4
Wild type n=6
CrebA n=8
E.
Area of secretory vesicles (µm) Single slice
0.006
0.005
0.004
0.003
0.002
0.001
0
**p<10-14
Wild type n=127
CrebA n=74
B.
Wild type
CrebA
C.
Wild type
CrebA

unexpected; whether or not this change is linked to CrebA secretory function remains to be determined.

To ask if CrebA is provided maternally and functions earlier in embryogenesis, we generated homozygous *CrebA* mutant germ-line clones using the FLP-DFS technique (Chou and Perrimon, 1996). Consistent with the absence of detectable germ-line expression (Rose et al., 1997; Smolik et al., 1992), maternal loss of *CrebA* did not exacerbate the *CrebA* zygotic loss-of-function phenotypes. Indeed, cuticle preparations of $CrebA^{mat\text{-}zyg\text{-}}$ animals revealed the same phenotypes as observed with zygotic mutants (Fig. 7A-D). Also, staining with the apical marker Crumbs, revealed that, as observed with only zygotic loss of *CrebA*, maternal-zygotic loss of *CrebA* did not result in overt defects in early embryos (Fig. 7E-H). Thus, CrebA is not required for the basal levels of secretion that occur in most cell types and instead functions to selectively upregulate secretory capacity in specialized secretory cells, both through its effects on genes encoding the general machinery and on genes encoding secreted cargo.

CrebA is related to the mammalian proteins Creb3L1 and Creb3L2

BLAST analysis of the Drosophila CrebA protein against the human genome revealed its two closest orthologues to be Creb3L1 (Honma et al., 1999; Nikaido et al., 2001) and Creb3L2 (Kondo et al., 2007; Saito et al., 2009) (Fig. 8A), consistent with earlier reports that Creb3L2, originally named BBF2H7, is most similar to Drosophila CrebA, originally named BBF-2 (Abel et al., 1992; Smolik et al., 1992; Kondo et al., 2007; Storlazzi et al., 2003). Creb3L1 and Creb3L2 share ~25% overall similarity with CrebA, with 97% similarity (84% identity) and 79% similarity (71% identity) within the DNA binding domains of CrebA with Creb3L1 and Creb3L2, respectively (Fig. 8B). Both human proteins also contain C-terminal transmembrane domains and are bound to the ER membrane (Kondo et al., 2007; Omori et al., 2002). During ER stress, Creb3L1 and Creb3L2

Figure 7. CrebA is not supplied maternally. (A-D) Cuticle preparations of larvae just prior to hatching from the vitelline membrane reveal that animals missing both the maternal (mat-) and zygotic (zyg-) contributions of *CrebA* are similar to animals missing only zygotic function (note the near complete absence of ventral denticles in B and C, and the loss of mouthpart pigmentation in the enlarged views, A' – D'). Paternally-supplied zygotic *CrebA*$^{+}$ is sufficient for maternal *CrebA* mutant larvae to finish embryonic development and hatch. (E-H) Staining with the Crumbs antibody reveals no overt morphological changes associated with either the maternal, zygotic or maternal and zygotic loss of CrebA. All embryos are early stage 15. All scale bars = 125μm.

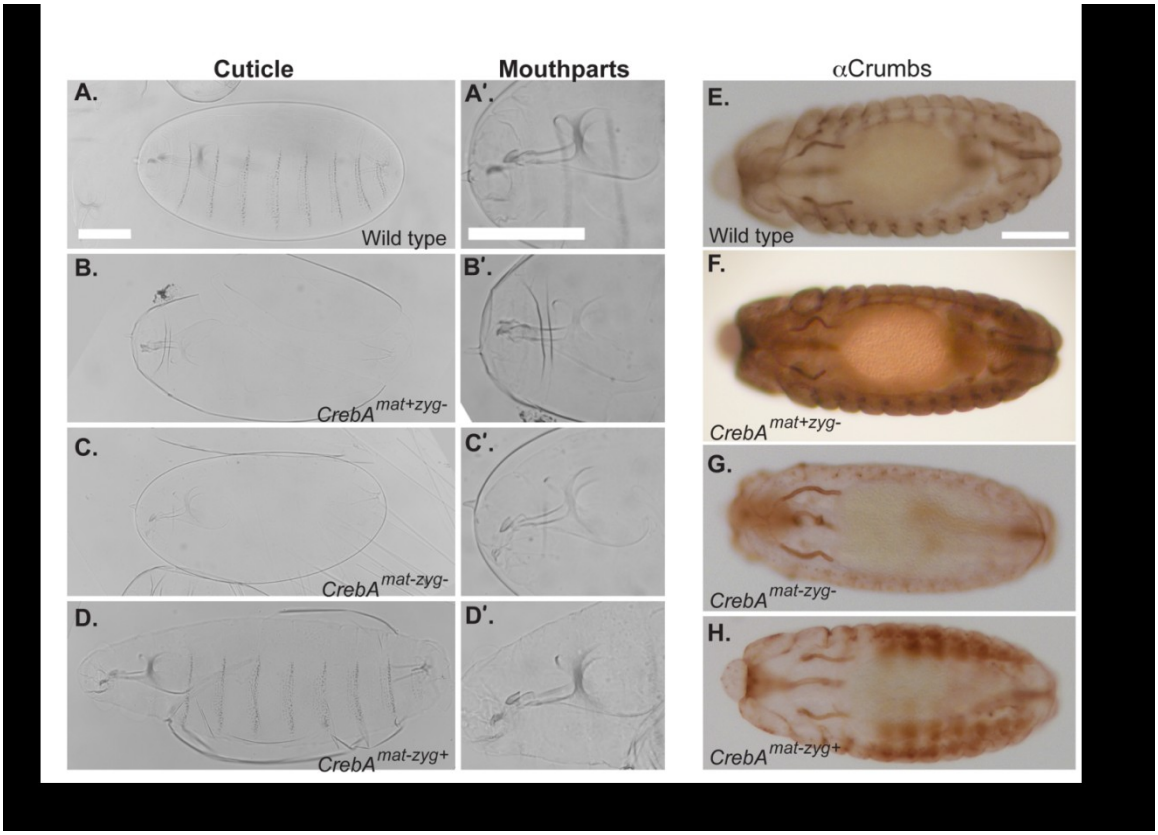
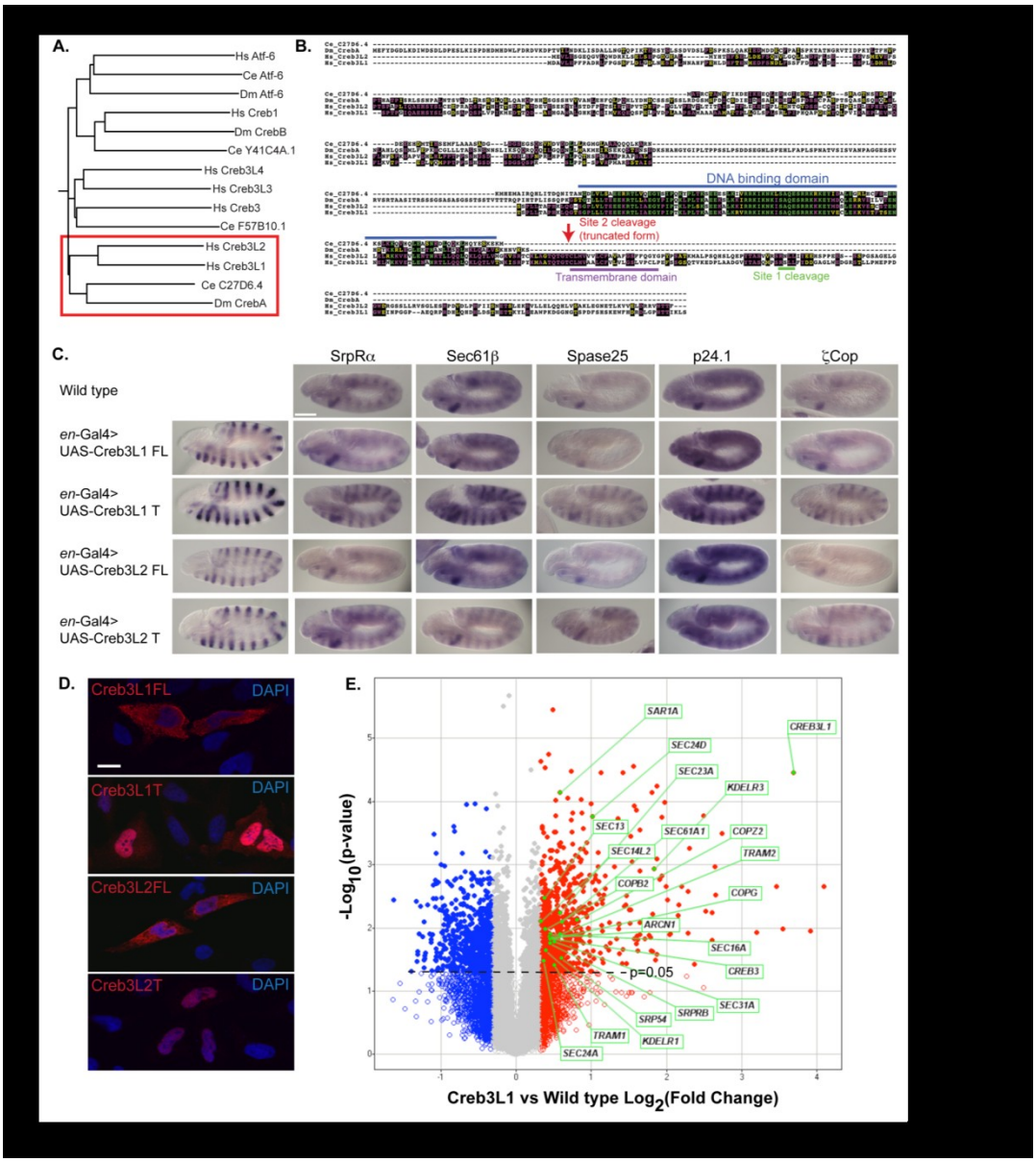

Cuticle
Mouthparts
αCrumbs
A.
A'.
E.
Wild type
Wild type
B.
B'.
F.
CrebA mat+zyg-
CrebA mat+zyg-
C.
C'.
G.
CrebA mat-zyg-
CrebA mat-zyg-
D.
D'.
H.
CrebA mat-zyg+
CrebA mat-zyg+

Figure 8. The CrebA human orthologues Creb3L1 and Creb3L2 are sufficient to upregulate the secretory pathway genes. (A) A rooted phylogenetic tree reveals the relationships among members of the CREB/ATF family of transcription factors from *Drosophila melanogaster* (Dm), *Caenorhabditis elegans* (Ce) and *Homo sapiens* (Hs). The red box highlights Drosophila CrebA and its closest family members. (B) Sequence alignment of the CrebA family. The truncated forms of Creb3L1 and Creb3L2 contain all N-terminal residues up to the site 2 protease cleavage point. (C) Human Creb3L1 and Creb3L2 upregulate SPCGs. The top row shows wild-type expression of each SPCG tested. Overexpression of UAS-Creb3L1 (FL or T), and UAS-Creb3L2 (FL or T) using the *en*-GAL4 driver resulted in robust mRNA expression in epidermal stripes (first column). Overexpression of the full-length forms did not affect SPCG expression, whereas expression of the active truncated forms induced high-level expression of every SPCG tested. Scale bar = 125□m. (D) HeLa cells transfected with either the full-length (FL) or truncated (T) forms of the Creb3L1 or Creb3L2 cDNAs. For both, the full-length protein is detected in the ER, whereas the active truncated form is predominantly detected in nuclei. Scale bar = 20□m. (E) Microarray analysis of HeLa cells expressing the active truncated form of Creb3L1 compared to control HeLa cells reveals a significant increase in the expression of many SPCGs. Genes represented in red (increased) or blue (decreased) show a fold change of at least 1.25. Open circles indicate non-significant changes.


A.
Hs Atf-6
Ce Atf-6
Dm Atf-6
Hs Creb1
Dm CrebB
Ce Y41C4A.1
Hs Creb3L4
Hs Creb3L3
Hs Creb3
Ce F57B10.1
Hs Creb3L2
Hs Creb3L1
Ce C27D6.4
Dm CrebA
B.
DNA binding domain
Site 2 cleavage
(truncated form)
Transmembrane domain
Site 1 cleavage
C.
SrpRα
Sec61β
Spase25
p24.1
ζCop
Wild type
en-Gal4>
UAS-Creb3L1 FL
en-Gal4>
UAS-Creb3L1 T
en-Gal4>
UAS-Creb3L2 FL
en-Gal4>
UAS-Creb3L2 T
D.
Creb3L1FL
DAPI
Creb3L1T
DAPI
Creb3L2FL
DAPI
Creb3L2T
DAPI
E.
SAR1A
CREB3L1
SEC24D
SEC23A
KDELR3
SEC13
SEC61A1
COPZ2
SEC14L2
TRAM2
COPB2
COPG
ARCN1
SEC16A
p=0.05
CREB3
SEC31A
SRPRB
SRP54
TRAM1
KDELR1
SEC24A
-Log10(p-value)
Creb3L1 vs Wild type Log2(Fold Change)

undergo regulated intramembrane proteolysis (RIP), wherein their transcription factor domains are released into the cytoplasm and subsequently translocated to the nucleus to activate transcription (Kondo et al., 2005; Kondo et al., 2007; Saito et al., 2007). The roles of both Creb3L1 and Creb3L2 as sensors and mediators in the unfolded protein response have been described and both are expressed in many secretory organs (Kondo et al., 2005; Kondo et al., 2007; Murakami et al., 2009; Omori et al., 2002; Saito et al., 2009; Saito et al., 2007).

Creb3L1 and Creb3L2 activate SPCG expression in Drosophila and human cells

To determine if human Creb3L1 and Creb3L2 have the same activity as CrebA, we generated UAS-lines that allow for Gal4 driven expression of both the full length (FL) and truncated (T) forms of Creb3L1 and Creb3L2 in Drosophila embryos. The truncated forms correspond to the fully processed proteins and should be functionally analogous to Drosophila CrebA (Kondo et al., 2005; Kondo et al., 2007). With the *en*-Gal4 driver, we detected high level expression of both forms of Creb3L1 and Creb3L2 in the epidermal stripes corresponding to *engrailed* expression (Fig. 8C). Importantly, we also detected robust expression of all five of the SPCGs we tested in the *engrailed* domain with expression of the truncated forms of both Creb3L1 and Creb3L2 but not with full-length Creb3L1 and Creb3L2, findings fully consistent with the robust localization of the truncated proteins to nuclei and of the full-length proteins to the ER (Kondo et al., 2005; Kondo et al., 2007). We conclude that human Creb3L1 and Creb3L2 can activate SPCG transcription in a heterologous system, suggesting a general and direct role for this family of bZip transcription factors in mediating high-level secretory capacity.

To ask if Creb3L1 can also upregulate SPCG expression in human cells, we carried out a microarray analysis of HeLa cells expressing the truncated form of Creb3L1 and control non-expressing cells. It should be noted that HeLa cells are not specialized for secretion, therefore

this experiment would determine if the overexpression of Creb3L1 alone can upregulate the secretory pathway in cells that are not dedicated to high-level protein secretion. Thus, it was exciting to discover that the most highly upregulated genes in Creb3L1 T expressing HeLa cells had GO terms including Golgi vesicle transport, secretory pathway, and secretion (Table S3, S4). The upregulated set of genes exhibited at least 3-fold enrichment in the prevalence of these terms as compared to the human genome (Table S4). Thus, like its Drosophila orthologue, Creb3L1 can activate expression of multiple components of the secretory pathway.

DISCUSSION

Here, we provide evidence that the CrebA/Creb3-like bZIP transcription factors are direct and major regulators of secretory capacity. Drosophila CrebA directly activates high-level expression of secretory pathway component genes through a site we found to be conserved among the enhancers of 34 CrebA-dependent SPCGs. Moreover, ectopic expression of CrebA in multiple tissues is sufficient to activate high-level expression of every SPCG tested. Microarray analysis indicates that CrebA is required for full expression of ~400 genes including almost 200 implicated in secretion. The secretory target genes include general machinery required for secretion in all cells as well as cell-type specific secreted cargo, such as the cuticle proteins and mucins. Phenotypic characterization of *CrebA* mutant SGs revealed a range of expected secretory defects, including reduced luminal secretory content and a decrease in the size and frequency of apical secretory vesicles, as well as unexpected changes in organelle distribution. We demonstrated that active forms of the closest vertebrate orthologues Creb3L1 and Creb3L2 activate the Drosophila SPCGs when expressed in embryos. Active Creb3L1 can also induce expression of multiple components of the secretory pathway when expressed in HeLa cells, a non-secretory cell type.

CrebA is the single Drosophila member of the Creb3-like family of transcription factors that includes five different proteins in mammals (Creb3/Luman, Creb3L1/Oasis, Creb3L2/BBF2H7, Creb3L3/CrebH and Creb3L4/Creb4) and two in worms (C27D6.4 and F57B10.1) (Fig. 7A). This singularity means that the fly protein is likely to play a more pivotal role in the regulation of secretion since there is no possibility of compensation for its activity by other family members. Each member of the Creb3-like family has a unique expression pattern with some overlap amongst family members. *Creb3/Luman* is most highly expressed in the brain, with expression detected in the liver, intestine, colon and skeletal muscles (Audas et al., 2008). *Creb3L1* is

expressed in osteoblasts, prostate, pancreas, ovary, testis, the gut, lungs, kidney and salivary glands (Nikaido et al., 2001; Omori et al., 2002). *Creb3L2* is expressed in chondrocytes, heart, lung, liver, kidney, adrenal gland, bladder, submandibular gland, brain, ovary, pancreas, spleen, testis and prostate (Kondo et al., 2007). *Creb3L3/CrebH* is almost exclusively detected in the liver (Chin et al., 2005), whereas *Creb3L4/Creb4* expression is elevated in the prostate, thymus, brain, pancreas, skeletal muscle and peripheral leukocytes (Cao et al., 2002). Unlike the Drosophila and worm orthologues, all five members of the Creb3-like family are ER-bound transcription factors previously implicated as sensors in the unfolded protein response (UPR) (Kondo et al., 2005; Kondo et al., 2007; Liang et al., 2006; Saito et al., 2007; Stirling and O'Hare, 2006; Zhang et al., 2006). Recently published phenotypes of the knock-out mutations in each of the two genes most closely related to *CrebA*, *Creb3L1* and Creb3L2, suggest a more physiological role for these genes during normal development, with a major defect being failure to secrete the extracellular matrix in the cell types expressing the highest levels of each gene (Murakami et al., 2009; Saito et al., 2009). These data support a model wherein one or more of the remaining members of the family may largely compensate for the loss of secretory capacity associated with the loss of any one family member. Indeed, our findings that the expression of only a single Creb3-like family member in HeLa cells, a non-secretory cell type, is sufficient to activate expression of multiple components of the secretory machinery further supports this hypothesis. Among the many secretory genes induced in HeLa cells by Creb3L1 are genes encoding multiple components of CopII vesicles: *Sec16A*, *Sec23A*, *Sec24A*, *Sec24D*, *Sec31A* and *Sar1A.* The reduced expression of one or more of these genes could explain the ER trapping of ECM proteins observed with the loss of either *Creb3L1* or *Creb3L2*.

Our microarray analysis of *CrebA* mutants revealed that CrebA upregulates transcription of secretory cargo, specifically expression of multiple components of the insect cuticle, several

mucin-like proteins (secreted highly-glycosylated proteins rich in serine and threonine), as well as multiple secreted proteins of unknown function. Although unexpected, this finding parallels the finding that mouse Creb3L1 directly upregulates the Type I collagen gene, *col1a1*, a major secreted component of bone ECM (Murakami et al., 2009). Our data also suggests that CrebA may function in parallel with tissue-specific regulators to control high-level expression of organ-specific cargo. An example is *CG14756*, which encodes an SG specific secreted protein of unknown function. Loss of *CrebA* results in a 3.2 fold decrease in the expression of this gene based on the microarray analysis, but unlike the CrebA targets that show more general expression in all secretory tissues, expression of *CG14756* could not be induced by CrebA in other cell types, suggesting the additional requirement for tissue-specific transcription factors for its activation (Fig. 4). Indeed, expression of *CG14756* is absolutely dependent on Fkh (R. Maruyama and D.J.A., unpublished), and the region immediately upstream of *CG14756* contains a good consensus Fkh binding site ~150 bp upstream of three clustered CrebA consensus binding sites. Thus, we propose that the CrebA/Creb3-like family enhances secretory capacity by coordinately upregulating expression of the general secretory machinery and of tissue-specific secreted cargo, with the expression of cargo genes likely mediated through cooperation with tissue-specific factors.

More than 30% (116 of 383) of genes identified in the CrebA microarray experiments had GO terms associated with roles in the secretory pathway (Fig. 4), and WoLF PSORT predictions suggested that more than half of the unknown targets are likely to have roles in secretion. Indeed, genes not implicated in the secretory pathway may, nonetheless, participate in secretion. Several of the ion channel/transporter genes have human orthologues known to function in secretory pathway organelles, as an example, *CG10449* (Drosophila *catsup*, human SLC39A7) encodes a Golgi localized zinc transporter (Huang et al., 2005). Also, 26 of the target

genes that did not have GO annotations have highly conserved human orthologues, several of which are involved in secretion. For example, *CG4293* and *CG7011* encode proteins similar to ERGIC2 and ERGIC3, respectively, proteins localized to the ER-Golgi intermediate compartment that function in protein folding and trafficking (Nishikawa et al., 2007). Thus, it is likely that many of the newly identified CrebA target genes encode proteins that function in secretory organelles, highlighting the potential of the microarray studies to reveal new genes with key roles in the efficient production and delivery of products through the secretory pathway.

Altogether, our studies reveal that CrebA and its human orthologues Creb3L1 and Creb3L2 activate transcription of components that function at all steps in secretion. Coordinate upregulation of secretory components by one (or a very few) transcription factors allows for easily adjustable levels of secretory capacity in a variety of cell types, as nicely exemplified in the Drosophila embryo where levels of CrebA and corresponding SPCG expression correlate with the levels of secretory activity in the different tissues. Furthermore, our microarray analysis combined with the recent studies of Creb3L1 and Creb3L2 in specialized cell types (osteoblasts and chrondrocytes) (Murakami et al., 2009; Saito et al., 2009) suggest that CrebA family proteins also upregulate expression of tissue-specific secreted content, highlighting the significance of this protein family in secretory cell specialization and function.

ACKNOWLEDGEMENTS

We would like to thank M. Anderson, D. Barrick, P. Devreotes, C. Machamer, and G. Seydoux for critical reading of the manuscript. In addition, we thank the members of the Andrew lab for comments and suggestions. We also thank T. Ruch and C. Machamer for assistance with the HeLa cell experiments, E. Chen and the Bloomington Stock Center for fly stocks, C. Talbot for assistance with microarray analysis, and C. Cooke for technical assistance with the TEM. This work is supported by NIH grants R01 DE013899 (D.J.A.) and F32 DE018853 (R.M.F.).

REFERENCES

Abel, T., R. Bhatt, and T. Maniatis. (1992). A Drosophila CREB/ATF transcriptional activator binds to both fat body- and liver-specific regulatory elements. *Genes Dev*. **6**, 466-80.

Abrams, E.W., and D.J. Andrew. (2005). CrebA regulates secretory activity in the Drosophila salivary gland and epidermis. *Development*. **132**, 2743-58.

Adachi, Y., K. Yamamoto, T. Okada, H. Yoshida, A. Harada, and K. Mori. (2008). ATF6 is a transcription factor specializing in the regulation of quality control proteins in the endoplasmic reticulum. *Cell Struct Funct*. **33**, 75-89.

Andrew, D.J., A. Baig, P. Bhanot, S.M. Smolik, and K.D. Henderson. (1997). The Drosophila dCREB-A gene is required for dorsal/ventral patterning of the larval cuticle. *Development*. *124*, 181-93.

Audas, T.E., Y. Li, G. Liang, and R. Lu. (2008). A novel protein, Luman/CREB3 recruitment factor, inhibits Luman activation of the unfolded protein response. *Mol Cell Biol*. *28*, 3952-66.

Birch-Machin, I., S. Gao, D. Huen, R. McGirr, R.A. White, and S. Russell. (2005). Genomic analysis of heat-shock factor targets in Drosophila. *Genome Biol*. **6**, R63.

Bommiasamy, H., S.H. Back, P. Fagone, K. Lee, S. Meshinchi, E. Vink, R. Sriburi, M. Frank, S. Jackowski, R.J. Kaufman, and J.W. Brewer. (2009). ATF6alpha induces XBP1-independent expansion of the endoplasmic reticulum. *J Cell Sci*. **122**, 1626-36.

Cao, G., X. Ni, M. Jiang, Y. Ma, H. Cheng, L. Guo, C. Ji, S. Gu, Y. Xie, and Y. Mao. (2002). Molecular cloning and characterization of a novel human cAMP response element-binding (CREB) gene (CREB4). *J Hum Genet*. **47**, 373-6.

Chin, K.T., H.J. Zhou, C.M. Wong, J.M. Lee, C.P. Chan, B.Q. Qiang, J.G. Yuan, I.O. Ng, and D.Y. Jin. (2005). The liver-enriched transcription factor CREB-H is a growth suppressor protein underexpressed in hepatocellular carcinoma. *Nucleic Acids Res*. **33**, 1859-73.

Chou, T.B., and N. Perrimon. (1996). The autosomal FLP-DFS technique for generating germline mosaics in Drosophila melanogaster. *Genetics*. **144**, 1673-9.

Dennis, G., Jr., B.T. Sherman, D.A. Hosack, J. Yang, W. Gao, H.C. Lane, and R.A. Lempicki. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*. *4*, P3.

Henderson, K.D., and D.J. Andrew. (2000). Regulation and function of Scr, exd, and hth in the Drosophila salivary gland. *Dev Biol*. **217**, 362-74.

Henderson, K.D., D.D. Isaac, and D.J. Andrew. (1999). Cell fate specification in the Drosophila salivary gland: the integration of homeotic gene function with the DPP signaling cascade. *Dev Biol*. **205**, 10-21.

Honma, Y., K. Kanazawa, T. Mori, Y. Tanno, M. Tojo, H. Kiyosawa, J. Takeda, T. Nikaido, T. Tsukamoto, S. Yokoya, and A. Wanaka. (1999). Identification of a novel gene, OASIS, which encodes for a putative CREB/ATF family transcription factor in the long-term cultured astrocytes and gliotic tissue. *Brain Res Mol Brain Res*. **69**, 93-103.

Huang da, W., B.T. Sherman, and R.A. Lempicki. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. **4**, 44-57.

Huang, L., C.P. Kirschke, Y. Zhang, and Y.Y. Yu. (2005). The ZIP7 gene (Slc39a7) encodes a zinc transporter involved in zinc homeostasis of the Golgi apparatus. *J Biol Chem*. **280**, 15456-63.

Irizarry, R.A., B.M. Bolstad, F. Collin, L.M. Cope, B. Hobbs, and T.P. Speed. (2003a). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. **31**, e15.

Irizarry, R.A., B. Hobbs, F. Collin, Y.D. Beazer-Barclay, K.J. Antonellis, U. Scherf, and T.P. Speed. (2003b). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. **4**, 249-64.

Kerman, B.E., A.M. Cheshire, and D.J. Andrew. (2006). From fate to function: the Drosophila trachea and salivary gland as models for tubulogenesis. *Differentiation*. **74**, 326-48.

Kondo, S., T. Murakami, K. Tatsumi, M. Ogata, S. Kanemoto, K. Otori, K. Iseki, A. Wanaka, and K. Imaizumi. (2005). OASIS, a CREB/ATF-family member, modulates UPR signalling in astrocytes. *Nat Cell Biol*. **7**, 186-94.

Kondo, S., A. Saito, S. Hino, T. Murakami, M. Ogata, S. Kanemoto, S. Nara, A. Yamashita, K. Yoshinaga, H. Hara, and K. Imaizumi. (2007). BBF2H7, a novel transmembrane bZIP transcription factor, is a new type of endoplasmic reticulum stress transducer. *Mol Cell Biol*. **27**, 1716-29.

Lee, A.H., G.C. Chu, N.N. Iwakoshi, and L.H. Glimcher. (2005). XBP-1 is required for biogenesis of cellular secretory machinery of exocrine glands. *Embo J*. **24**, 4368-80.

Lehmann, R., and D. Tautz. (1994). In situ hybridization to RNA. *Methods Cell Biol*. **44**, 575-98.

Liang, G., T.E. Audas, Y. Li, G.P. Cockram, J.D. Dean, A.C. Martyn, K. Kokame, and R. Lu. (2006). Luman/CREB3 induces transcription of the endoplasmic reticulum (ER) stress response protein Herp through an ER stress response element. *Mol Cell Biol*. **26**, 7999-8010.

Livak, K.J., and T.D. Schmittgen. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. **25**, 402-8.

Montminy, M.R., and L.M. Bilezikjian. (1987). Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene. *Nature*. **328**, 175-8.

Murakami, T., A. Saito, S. Hino, S. Kondo, S. Kanemoto, K. Chihara, H. Sekiya, K. Tsumagari, K. Ochiai, K. Yoshinaga, M. Saitoh, R. Nishimura, T. Yoneda, I. Kou, T. Furuichi, S. Ikegawa, M. Ikawa, M. Okabe, A. Wanaka, and K. Imaizumi. (2009). Signalling mediated by the endoplasmic reticulum stress transducer OASIS is involved in bone formation. *Nat Cell Biol*. **11**, 1205-11.

Myat, M.M., D.D. Isaac, and D.J. Andrew. (2000). Early genes required for salivary gland fate determination and morphogenesis in Drosophila melanogaster. *Adv Dent Res*. **14**, 89-98.

Nikaido, T., S. Yokoya, T. Mori, S. Hagino, K. Iseki, Y. Zhang, M. Takeuchi, H. Takaki, S. Kikuchi, and A. Wanaka. (2001). Expression of the novel transcription factor OASIS, which belongs to the CREB/ATF family, in mouse embryo with special reference to bone development. *Histochem Cell Biol*. **116**, 141-8.

Nishikawa, M., Y. Kira, Y. Yabunaka, and M. Inoue. (2007). Identification and characterization of endoplasmic reticulum-associated protein, ERp43. *Gene*. 386, 42-51.

Omori, Y., J. Imai, Y. Suzuki, S. Watanabe, A. Tanigami, and S. Sugano. (2002). OASIS is a transcriptional activator of CREB/ATF family with a transmembrane domain. *Biochem Biophys Res Commun*. **293**, 470-7.

Reuter, R., G.E. Panganiban, F.M. Hoffmann, and M.P. Scott. (1990). Homeotic genes regulate the spatial expression of putative growth factors in the visceral mesoderm of Drosophila embryos. *Development*. 110, 1031-40.

R**ose, R.E., N.M. Gallaher, D.J. Andrew, R.H. Goodman, and S.M. Smolik.** (1997). The CRE-binding protein dCREB-A is required for Drosophila embryonic development. *Genetics*. **146**, 595-606.

Saito, A., S. Hino, T. Murakami, S. Kanemoto, S. Kondo, M. Saitoh, R. Nishimura, T. Yoneda, T. Furuichi, S. Ikegawa, M. Ikawa, M. Okabe, and K. Imaizumi. (2009). Regulation of endoplasmic reticulum stress response by a BBF2H7-mediated Sec23a pathway is essential for chondrogenesis. *Nat Cell Biol*. **11**, 1197-204.

Saito, A., S. Hino, T. Murakami, S. Kondo, and K. Imaizumi. (2007). A novel ER stress transducer, OASIS, expressed in astrocytes. *Antioxid Redox Signal*. **9**, 563-71.

Sbodio, J.I., and C.E. Machamer. (2007). Identification of a redox-sensitive cysteine in GCP60 that regulates its interaction with golgin-160. *J Biol Chem*. **282**, 29874-81.

Shaffer, A.L., K.I. Lin, T.C. Kuo, X. Yu, E.M. Hurt, A. Rosenwald, J.M. Giltnane, L. Yang, H. Zhao, K. Calame, and L.M. Staudt. (2002). Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity*. **17**, 51-62.

Shaffer, A.L., M. Shapiro-Shelef, N.N. Iwakoshi, A.H. Lee, S.B. Qian, H. Zhao, X. Yu, L. Yang, B.K. Tan, A. Rosenwald, E.M. Hurt, E. Petroulakis, N. Sonenberg, J.W. Yewdell, K. Calame, L.H. Glimcher, and L.M. Staudt. (2004). XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. *Immunity*. **21**, 81-93.

Shiga, Y., Tanaka-Matakatsu, M. and Hayashi, S. (1996). A nuclear GFP/b-galactosidase fusion protein as a marker for morphogenesis in living *Drosophila*. *Development, Growth & Differentiation*. **38**, 99-106.

Smolik, S.M., R.E. Rose, and R.H. Goodman. (1992). A cyclic AMP-responsive element-binding transcriptional activator in Drosophila melanogaster, dCREB-A, is a member of the leucine zipper family. *Mol Cell Biol*. **12**, 4123-31.

Stirling, J., and P. O'Hare. (2006). CREB4, a transmembrane bZip transcription factor and potential new substrate for regulation and cleavage by S1P. *Mol Biol Cell*. **17**, 413-26.

Storlazzi, C.T., F. Mertens, A. Nascimento, M. Isaksson, J. Wejde, O. Brosjo, N. Mandahl, and I. Panagopoulos. (2003). Fusion of the FUS and BBF2H7 genes in low grade fibromyxoid sarcoma. *Hum Mol Genet*. **12**, 2349-58.

Wang, Y., J. Shen, N. Arenzana, W. Tirasophon, R.J. Kaufman, and R. Prywes. (2000). Activation of ATF6 and an ATF6 DNA binding site by the endoplasmic reticulum stress response. *J Biol Chem*. **275**, 27013-20.

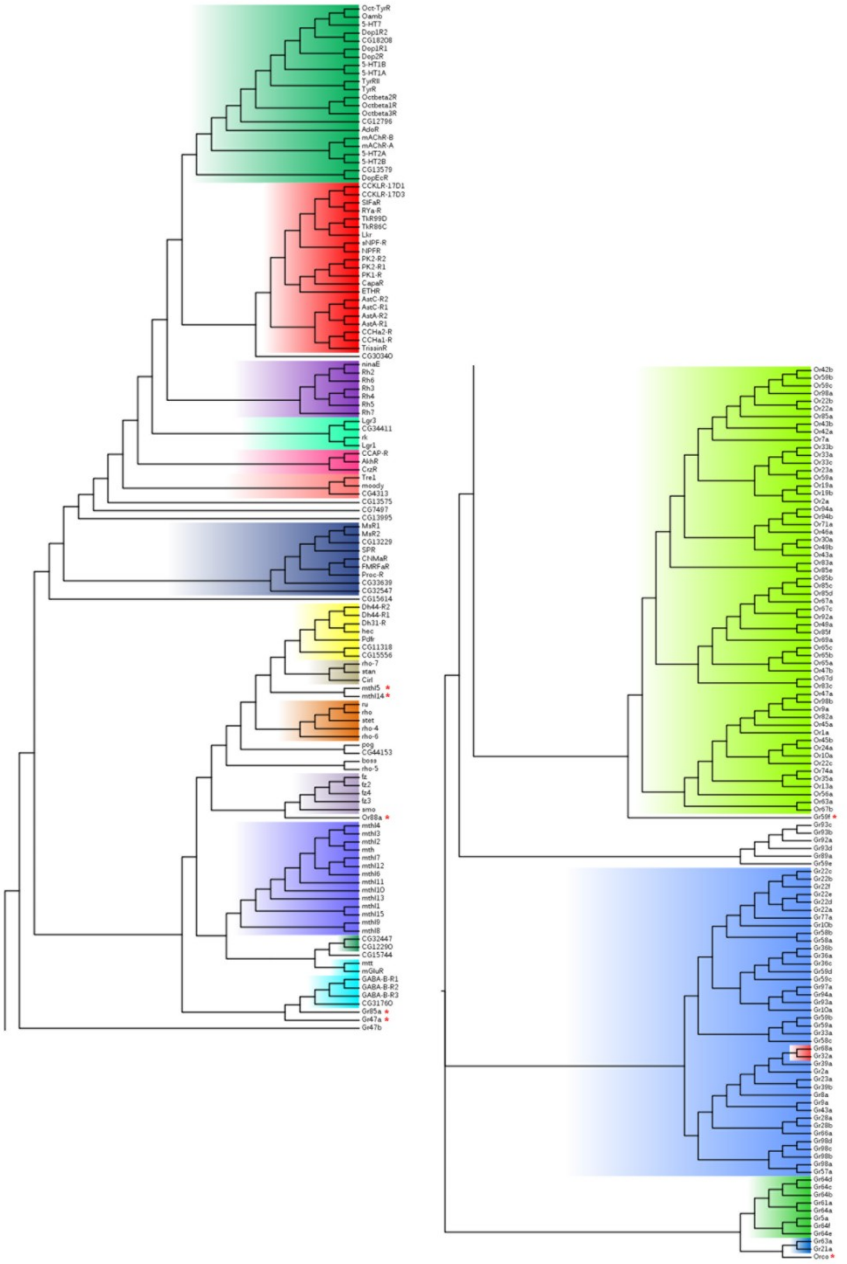
Weiss, J.B., K.L. Suyama, H.H. Lee, and M.P. Scott. (2001). Jelly belly: a Drosophila LDL receptor repeat-containing signal required for mesoderm migration and differentiation. *Cell*. **107**, 387-98.

Zhang, K., X. Shen, J. Wu, K. Sakaki, T. Saunders, D.T. Rutkowski, S.H. Back, and R.J. Kaufman. (2006). Endoplasmic reticulum stress activates cleavage of CREBH to induce a systemic inflammatory response. *Cell*. **124**, 587-99.

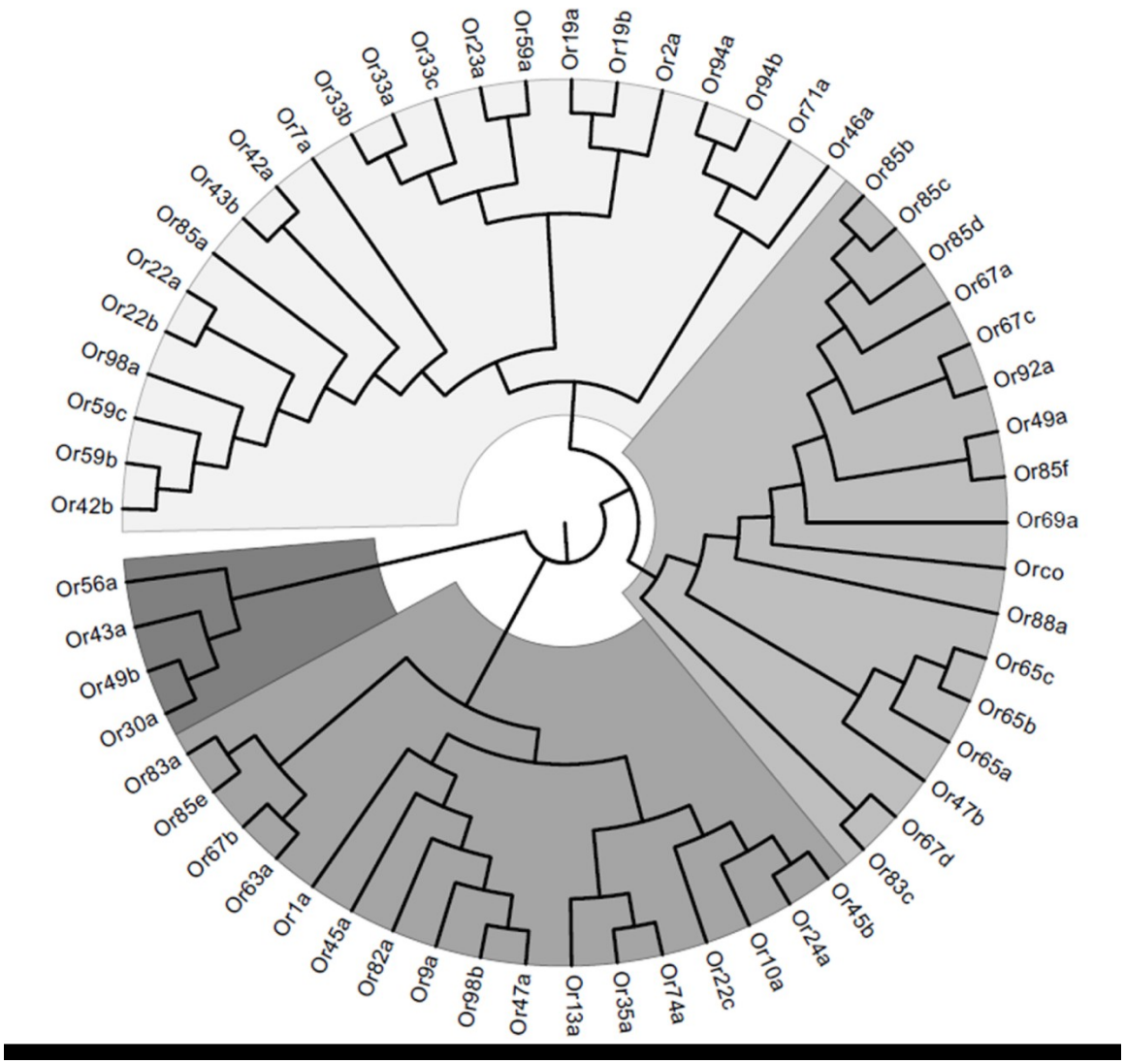
APPENDIX A:

Supplementary Materials from Chapter 2

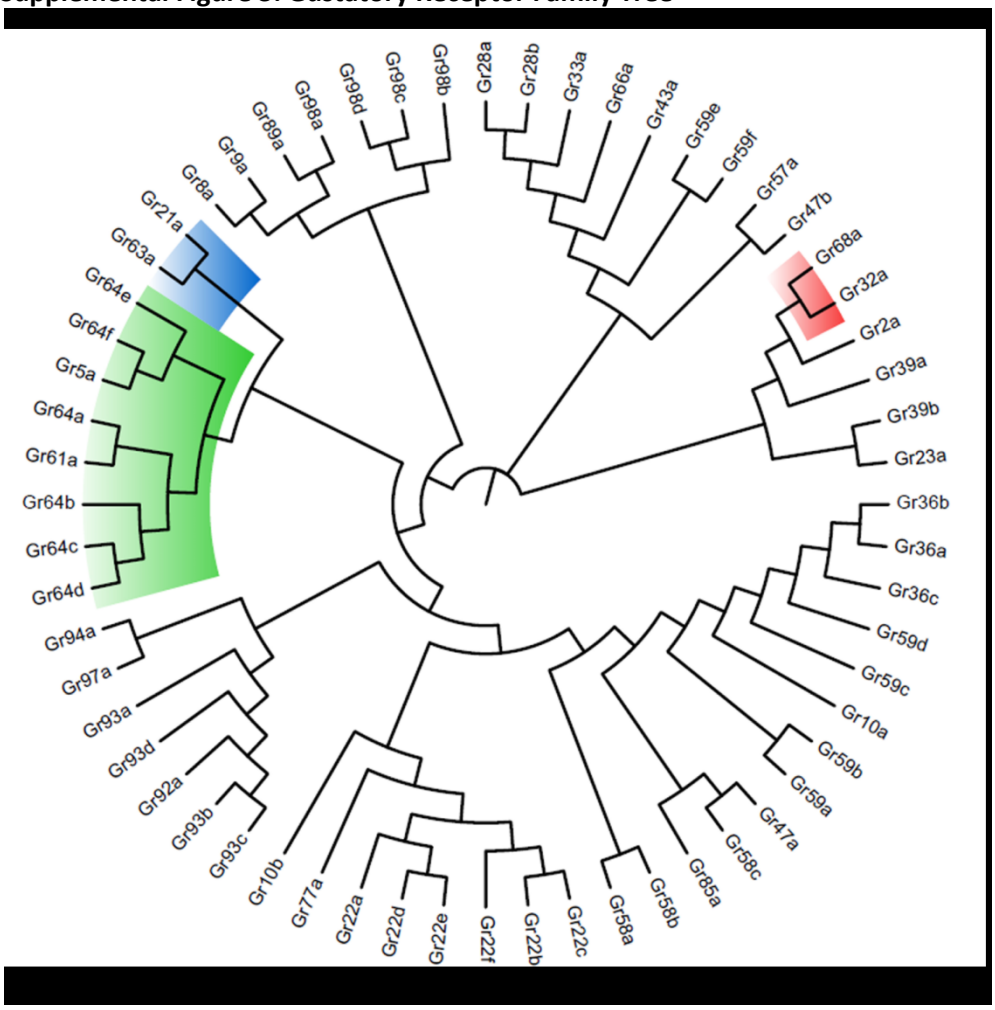
Supplemental Figure 1: Phylogentic Tree of all GPCRs encoded in *Drosophila*



Supplemental Figure 2: Odorant Receptor Family Tree



Supplemental Figure 3: Gustatory Receptor Family Tree



Supplemental Table 1: Names and Signaling Information for Drosophila NOG GPCRs

Name	Symbol	CG	Output	G-alpha	Assay	ligand	*	**
GABA-B receptor subtype 3	GABA-B-R3	CG3022					C1	31
GABA-B receptor subtype 1	GABA-B-R1	CG15274	↑ Ca2+, ↓ cAMP (Mezler 2001)		HEK293 cells, Xenopus	GABA	C1	31
GABA-B receptor subtype 2	GABA-B-R2	CG6706	↑ Ca2+, ↓ cAMP (Mezler 2001)		HEK293 cells, Xenopus	GABA	C1	31
mangetout	mtt	CG30361	↑ IP via Gαi, Gαo (Mitri 2004)		HEK293 cells	L-canavanine		28
Glutamate Receptor	mGluR	CG11144	PTX sensitivity (Raymond 1999)		Xenopus	glutamate		
poor gastrulation	pog	CG31660						31
CG31760	CG31760	CG31760						31
CG44153	CG44153	CG44153						
CG12290	CG12290	CG12290					A5b	24
CG32447	CG32447	CG32447						31
CG15744	CG15744	CG15744						
frizzled	fz	CG17697	Gαs (NiCHO cellsls 2013)	Gαs	genetic interaction	Wnt	D1	32
frizzled 2	fz2	CG9739				Wnt	D1	32
frizzled 3	fz3	CG16785				Wnt	D1	32
frizzled 4	fz4	CG4626				Wnt	D1	32
smoothened	smo	CG11561	↓ cAMP (Ogden 2008)	Gαi	Cl8 cells, genetic interaction		D1	32
Ca-independent α-latrotoxin	Cirl	CG8639					B4	27
starry night	stan	CG11895					D2	27
CG11318	CG11318	CG11318					B3	27
CG15556	CG15556	CG15556					B3	26
hector	hec	CG4395					B1	UN
Diuretic hormone 31	Dh31-R	CG32843					B2	25
Diuretic hormone 44 receptor 2	Dh44-R2	CG12370	↑ cAMP (Hector 2009)		HEK293 cells	Diuretic hormone 44	B2	25
Diuretic hormone 44 receptor 1	Dh44-R1	CG8422	↑ Ca2+, ↑ cAMP (Johnson 2004)		HEK293 cells	Diuretic hormone 44	B2	25
Pdfr	CG13758	CG13758	↑ cAMP, ↑ Ca2+ (Hyun 2005, Agrawal 2013)		S2 cells, RNAi screen	pigment-dispersing factor	B1	25
methuselah-like 4	mthl4	CG6536					B5	26
methuselah-like 3	mthl3	CG6530					B5	26
methuselah-like 2	mthl2	CG17795					B5	26
methuselah	mth	CG6936	↑ Ca2+ (Cvejic 2004)		HEK293 cells	Sun, SP, SPAM	B5	26
methuselah-like 12	mthl12	CG32853					B5	26
methuselah-like 7	mthl7	CG7476					B5	26
methuselah-like 6	mthl6	CG16992					B5	26
methuselah-like 11	mthl11	CG31147					B5	26
methuselah-like 13	mthl13	CG30018					B5	26
methuselah-like 10	mthl10	CG17061					B5	26
methuselah-like 15	mthl15	CG31720						26
Mist	mthl1	CG4521	G12/13 (Manning 2013)	cta	S2 cells	Fog		26
methuselah-like 5	mthl5	CG6965						26
methuselah-like 14	mthl14	CG32476						26
methuselah-like 9	mthl9	CG17084					B5	26
methuselah-like 8	mthl8	CG32475					B5	26
bride of sevenless	boss	CG8285					D2	30
roughoid	ru	CG1214						
rhomboid	rho	CG1004						
stem cell tumor	stet	CG33166						
rhomboid-4	rho-4	CG1697						
rhomboid-6	rho-6	CG17212						
rhomboid-5	rho-5	CG33304						
rhomboid-7	rho-7	CG8972						
CG32547	CG32547	CG32547						
CG4313	CG4313	CG4313						21
moody	moody	CG4322	Gαi, Gαo (Schwabe 2005)	Gαi, Gαo	localization			21
Trapped in endoderm 1	Tre1	CG3171	Gαi, Gαo (Yoshiura 2012)	Gαi, Gαo	physical & genetic interaction		A5a	
CG7497	CG7497	CG7497					A5b	18
CG13995	CG13995	CG13995					A5b	23
Myosuppressin receptor 1	MsR1	CG8985	↓ cAMP (Johnson 2003)		HEK293 cells	dromyosupressin	A5a	5
Myosuppressin receptor 2	MsR2	CG43745	↓ cAMP (Johnson 2003)		HEK293 cells	dromyosupressin	A5a	
CG13299	CG13299	CG13229					A5a	5
Sex peptide receptor	SPR	CG16752	↑ Ca2+ (Yapici 2008), Gαq, Gαi, Gαo		CHO cells	SP		5
CG33639	CG33639	CG33639					A5a	UN
FMRFamide Receptor	FMRFaR	CG2114	↑ Ca2+ (Cazzamali 2002)		CHO cells	FMRFamides	A5a	
Proctolin receptor	Proc-R	CG6986	↑ Ca2+ (Egerod 2003)		CHO cells	proctolin	A5b	5
CNMamide Receptor	CNMaR	CG33696	↑ Ca2+ (Jung 2014)		CHO cells	CNMa		5
CG15614	CG15614	CG15614						
Leucine-rich repeat-GPCR3	Lgr3	CG31096					A4b	7
CG34411	CG34411	CG34411					A4b	7
rk	rickets	CG8930	↑ cAMP (Luo 2005)		293T cells	burs and pburs	A4b	7
Leucine-rich GPCR1	Lgr1	CG7665	↑ cAMP (Sudo 2005)		293T cells	GPA2/GPB5		7
CG13575	CG13575	CG13575					A4i	11
Tachykinin-like receptor at 99D	TkR99D	CG7887	↑ Ca2+ (Poels 2007)		S2 cells	tachykinins	A4f	11
Tachykinin-like receptor at 86C	TkR86C	CG6515	↑ IP (Monnier 1992)		NIH 3T3 cells	tachykinin	A4f	11
Leucokinin receptor	Lkr	CG10626	↑ Ca2+ (Radford 2002)		S2 cells	Drosokinin	A4f	11
RYamide receptor	RYa-R	CG5811	↑ Ca2+ (Ida 2011)		CHO cells	dRYamide-1 and dRYamide-2		
CHO cellslecystokinin-like 17D1	CCKLR-17D1	CG42301	↑ cAMP (Chen 2012)		genetic	dsk-1, 2		8
CHO cellslecystokinin-like 17D3	CCKLR-17D3	CG32540	↑ Ca2+, PTX sensitive (Kubiak 2002)		CHO cells, HEK293, SH-EP	dsk-1		8
SIFamide receptor	SIFaR	CG10823	↑ Ca2+ (Jorgensen 2006)		CHO cells	SIFamide	A4i	
short neuropeptide F receptor	sNPF-R	CG7395	↑ Ca2+ (Mertens 2002); ↓ cAMP (Vecsey 2014)	Gαo	CHO cells	RX_1RX_2 amide C-term short NPFs		11
Neuropeptide F receptor	NPFR	CG1147	↓ cAMP (Garczynkski 2002)		CHO cells	NPF	A4f	11
Pyrokinin 2 receptor 2	PK2-R2	CG8795	↑ Ca2+ (Rosenkilde 2003)		CHO cells	pyrokinin-2	A4e	5
Pyrokinin 2 receptor 1	PK2-R1	CG8784	↑ Ca2+ (Rosenkilde 2003)		CHO cells	pyrokinin-2	A4e	5
Pyrokinin 1 receptor	PK1-R	CG9918	↑ Ca2+ (Cazzamali 2005)		CHO cells	pyrokinin-1	A4e	5
Capability receptor	CapaR	CG14575	↑ Ca2+ (Park 2002)		Xenopus oocyte	CAPZb-1	A4i	5
ETHR	ETHR	CG5911	↑ Ca2+ (Park 2003)		CHO cells	ETH1, MasETH, MasPETH	A4i	5
Allatostatin C receptor 2	AstC-R2	CG13702					A4g	1
Allatostatin C receptor 1	AstC-R1	CG7285					A4g	1
Allatostatin A receptor 2	AstA-R2	CG10001	↑ Ca2+, PTX sensitive (Larsen 2001)		CHO cells	DST-As	A4a	2
Allatostatin A receptor 1	AstA-R1	CG2872	↑ Ca2+, PTX sensitive (Larsen 2001)		CHO cells	DST-As		2
CCHamide-2 receptor	CCHa2-R	CG14593	↑ Ca2+ (Hansen 2011)		CHO cells	CCHamide-1, 2		4
CCHamide-1 receptor	CCHa1-R	CG30106	↑ Ca2+ (Hansen 2011)		CHO cells	CCHamide-1, 2		4
Trissin receptor	TrissinR	CG34381	↑ Ca2+ (Ida 2011)		CHO cells	Trissin	A4i	
CG30340	CG30340	CG30340						8

ninal E	ninaE	CG4550	↑ PLC activity (Running Deer 1995)	Gαq	Drosophila extract	light	A1	14
Rhodopsin 2	Rh2	CG16740	↑ Ca2+ (Lee 1990; Cook 1999)	Gαq	localization	light	A1	14
Rhodopsin 6	Rh6	CG5192	↑ Ca2+ (Lee 1990; Cook 1999)	Gαq	localization	light	A1	14
Rhodopsin 3	Rh3	CG10888	↑ Ca2+ (Lee 1990; Cook 1999)	Gαq	localization	light	A1	14
Rhodopsin 4	Rh4	CG9668	↑ Ca2+ (Lee 1990; Cook 1999)	Gαq	localization	light	A1	14
Rhodopsin 5	Rh5	CG5279	↑ Ca2+ (Lee 1990; Cook 1999)	Gαq	localization	light	A1	14
Rhodopsin 7	Rh7	CG5638	↑ Ca2+ (Lee 1990; Cook 1999)	Gαq	localization	light	A1	14
Dopamine 1-like receptor 2	Dop1R2	CG18741	↑ Ca2+, ↑ cAMP (Feng 1996)		Xenopus oocytes	dopamine	A2b	24
Octopamine in mushroom bodies	Oamb	CG3856	↑ Ca2+, ↑ cAMP (Han 1998)		HEK293, S2 cells	octopamine	A2d	24
CG18208	CG18208	CG18208						24
Dopamine 1-like receptor 1	Dop1R1	CG9652	↑cAMP (Gotzes 1994)		HEK293 cells	dopamine	A2b	24
Dopamine 2-like receptor	Dop2R	CG33517	↓ cAMP, PTX sensitive (Hearn 2002)		HEK293 cells	dopamine		24
5-hydroxytryptamine 7	5-HT7	CG12073	↑ cAMP (Saudaou 1992)		NIH 3T3 cells	serotonin and more	A2a	22
Octopamine-Tyramine receptor	Oct-TyrR	CG7485	↓ cAMP, ↑ Ca2+ (Robb 1994)		CHO cells	octopamin, tyramine		24
5-hydroxytryptamine 1B	5-HT1B	CG15113	↑ IP; ↓ cAMP (Saudou 1992)		NIH 3T3 cells	serotonin	A2a	22
5-hydroxytryptamine 1A	5-HT1A	CG16720	↑ IP; ↓ cAMP (Saudou 1992)		NIH 3T3 cells	serotonin	A2a	22
Tyramine receptor II	TyrRII	CG16766	↑ cAMP (Bayliss 2013)		CHO cells	octopamine, dopamine	A2e	22
Tyramine receptor	TyrR	CG7431	↑ Ca2+ (Cazzamali 2005)		CHO cells	tyramine	A2e	22
Octopamine β2 receptor	Octβ2R	CG33976	↑ cAMP (Maqueira 2005)		CHO cells	octopamine		20
Octopamine β1 receptor	Octβ1R	CG6919	↑ cAMP (Balfanz 2005)		HEK293, Flp-In-293	octopamine	A2e	20
Octopamine β3 receptor	Octβ3R	CG42244	↑ cAMP (Maqueira 2005)		CHO cells	octopamine		
5-hydroxytryptamine 2A	5-HT2a	CG1056					A2a	24
5-hydroxytryptamine 2B	5-HT2B	CG42796					A2e	24
muscarinic AcetylCHO, A-type	mAChR-A	CG4356	↑ Ca2+ (Millar 1995)		S2 cells	carbamylCHO,acetylCHO, muscarine	A2c	23
muscarinic AcetylCHO, B-type	mAChR-B	CG7918	↑ Ca2+ (Collin 2013)		CHO cells		A2c	23
CG13579	CG13579	CG13579					A5b	UN
Dopamine/Ecdysteroid receptor	DopEcR	CG18314	↑ cAMP (Srivastava 2005)		CHO cells	20-hydroxyecdysone, dopamine	A2e	UN
Crustacean cardioactive peptide	CCAP-R	CG33344	↑ Ca2+ (Cazzamali 2003)		CHO cells	crustacean cardioactive peptide		6
AkhR	AkhR	CG11325	↑ Ca2+ (Park 2002)		Xenopus oocyte	AKH		
CrzR	CrzR	CG10698	↑ Ca2+ (Park 2002)		Xenopus oocyte	corazonin	A4d	6
Adenosine receptor	AdoR	CG9753	↑ cAMP, ↑ Ca2+ (Dolezelova 2007)	Gas	CHO cells		A3	23
CG12796	CG12796	CG12796					A2e	24

Supplemental Table 2: Odorant receptor family and information

Name	Symbol	CG	G-alpha/Output	ligand
Odorant receptor 42b	Or42b	CG12754		ethyl acetate and pentyl acetate
Odorant receptor 59b	Or59b	CG3569		DEET, methyl acetate
Odorant receptor 59c	Or59c	CG17226		many
Odorant receptor 98a	Or98a	CG5540		many
Odorant receptor 22b	Or22b	CG4231		esters
Odorant receptor 22a	Or22a	CG12193	↑ Ca2+ via Gαi/Gαo (Raja 2014)	ethyl butyrate, methyl hexanoate, esters, and others
Odorant receptor 85a	Or85a	CG7454		ethyl 3-hydroxybutyrate
Odorant receptor 43b	Or43b	CG17853		ethyl butyrate
Odorant receptor 42a	Or42a	CG17250		butanol, ethyl acetate, propyl acetate, pentyl acetate, pyrazines
Odorant receptor 7a	Or7a	CG10759		E2 hexenal
Odorant receptor 33b	Or33b	CG16961		pentyl acetate and pyrazines
Odorant receptor 33a	Or33a	CG16960		many
Odorant receptor 33c	Or33c	CG5006		ethyl acetate, cyclohexanone, fenchone
Odorant receptor 23a	Or23a	CG9880		many
Odorant receptor 59a	Or59a	CG9820		ethyl acetate, anisole, hexanoic acid, and pyrazines
Odorant receptor 19a	Or19a	CG18859		valencene, 1-octen-3-ol
Odorant receptor 19b	Or19b	CG32825		many
Odorant receptor 2a	Or2a	CG3206		many
Odorant receptor 94a	Or94a	CG17241		many
Odorant receptor 94b	Or94b	CG6679		many
Odorant receptor 71a	Or71a	CG17871		4-methyl phenol
Odorant receptor 46a	Or46a	CG33478		4-methyl phenol
Odorant receptor 85b	Or85b	CG11735		2-heptanone, amyl acetate, and butyl acetate
Odorant receptor 85c	Or85b	CG17911		many
Odorant receptor 85d	Or85d	CG11742		isoamyl acetate, 2-heptanone
Odorant receptor 67a	Or67a	CG12526		benzaldehyde and acetophenone
Odorant receptor 67c	Or67c	CG14156		ethyl lactate
Odorant receptor 92a	Or92a	CG17916		many
Odorant receptor 49a	Or49a	CG13158		butanol and 2-heptanone
Odorant receptor 85f	Or85f	CG16755		many
Odorant receptor 69a	Or69a	CG33264		many
Odorant receptor co-receptor	Orco	CG10609		many
Odorant receptor 88a	Or88a	CG14360		male and female pheromones
Odorant receptor 65c	Or65c	CG32403		many
Odorant receptor 65b	Or65b	CG32402		many
Odorant receptor 65a	Or65a	CG32401		male pheromone (11-cis-vaccenyl acetate)
Odorant receptor 47b	Or47b	CG13206		male and female pheromones
Odorant receptor 67d	Or67d	CG14157		male pheromone (11-cis-vaccenyl acetate)
Odorant receptor 83c	Or83c	CG15581		farnesol and others
Odorant receptor 45b	Or45b	CG12931		anisole
Odorant receptor 24a	Or24a	CG11767		pyrazines, pentanol, hexanol, octanol, and many others
Odorant receptor 10a	Or10a	CG17867		ethyl hexanoate, ethyl benzoate, benzaldehyde, and acetophenone
Odorant receptor 22c	Or22c	CG15377		hexanol, pentyl acetate, benzyl acetate, and 2-heptanone
Odorant receptor 74a	Or74a	CG13726		octanol, anisole, and 2-heptanone
Odorant receptor 35a	Or35a	CG17868		butanol, pentanol, hexanol, octanol, propyl acetate, butyl acetate, esters
Odorant receptor 13a	Or13a	CG12697		octanol, nonanol, and pentyl acetate
Odorant receptor 47a	Or47a	CG13225		pentyl acetate
Odorant receptor 98b	Or98b	CG1867		ethyl benzoate
Odorant receptor 9a	Or9a	CG15302		2-pentanol
Odorant receptor 82a	Or82a	CG31519		geranyl acetate
Odorant receptor 45a	Or45a	CG1978		hexanol, pentyl acetate, benzyl acetate, and 2-heptanone
Odorant receptor 1a	Or1a	CG17885		butanal, heptanal, and 2-heptanone
Odorant receptor 63a	Or63a	CG9969		butyl acetate, isoamyl acetate, and hexanoic acid
Odorant receptor 67b	Or67b	CG14176		ethyl acetate, pentyl acetate, methyl caproate, and many others
Oderant receptor 85e	Or85e	CG9700		ethyl acetate, cyclohexanone, fenchone
Odorant receptor 83a	Or83a	CG10612		pentanol, ethyl acetate, and propyl acetate
Odorant receptor 30a	Or30a	CG13106		propyl acetate and anisole
Odorant receptor 49b	Or49b	CG17584		2-mthylphenol
Odorant receptor 43a	Or43a	CG1854		acetophenone, benzaldehyde, benzyl alcohol, hexanol, cyclohexanone, and others
Odorant receptor 56a	Or56a	CG12501		geosmin

Supplemental Table 3: Gustatory receptor family and information

Name	Symbol	CG	G-alpha/Output	ligand	Assay
Gustatory receptor 36b	Gr36b	CG31744			
Gustatory receptor 36a	Gr36a	CG31747			
Gustatory receptor 36c	Gr36c	CG31748			
Gustatory receptor 59d	Gr59d	CG30330			
Gustatory receptor 59c	Gr59c	CG30186			
Gustatory receptor 10a	Gr10a	CG32664			
Gustatory receptor 59b	Gr59b	CG30191			
Gustatory receptor 59a	Gr59a	CG30189			
Gustatory receptor 47a	Gr47a	CG12906			
Gustatory receptor 58c	Gr58c	CG13491			
Gustatory receptor 85a	Gr85a	CG31405			
Gustatory receptor 58b	Gr58b	CG13495			
Gustatory receptor 58a	Gr58a	CG30396			
Gustatory receptor 22c	Gr22c	CG31929			
Gustatory receptor 22b	Gr22b	CG31931		bitter	
Gustatory receptor 22f	Gr22f	CG31932			
Gustatory receptor 22e	Gr22e	CG31936			
Gustatory receptor 22d	Gr22d	CG31930			
Gustatory receptor 22a	Gr22a	CG31662			
Gustatory receptor 77a	Gr77a	CG32433			
Gustatory receptor 10b	Gr10b	CG12622			
Gustatory receptor 93c	Gr93c	CG31173			
Gustatory receptor 93b	Gr93b	CG31336			
Gustatory receptor 92a	Gr92a	CG31208			
Gustatory receptor 93d	Gr93d	CG31335			
Gustatory receptor 93a	Gr93a	CG13417		bitter, caffiene	
Gustatory receptor 97a	Gr97a	CG33083			
Gustatory receptor 94a	Gr94a	CG31280			
Gustatory receptor 64d	Gr64d	CG14987		sweet taste	
Gustatory receptor 64c	Gr64c	CG32256		sweet taste	
Gustatory receptor 64b	Gr64b	CG32257		sweet taste	
Gustatory receptor 61a	Gr61a	CG13888		glucose	
Gustatory receptor 64a	Gr64a	CG32261		maltose, sucrose	
Gustatory receptor 5a	Gr5a	CG15779	Gαs (Ueno 2006)	trehalose	genetics
Gustatory receptor 64f	Gr64f	CG32255		sucrose, maltose, and glucose	
Gustatory receptor 64e	Gr64e	CG32258		sweet taste	
Gustatory receptor 63a	Gr63a	CG14979	Gαq (Yao 2010)	CO2	in vivo recordings
Gustatory receptor 21a	Gr21a	CG13948	Gαq (Yao 2010)	CO2	in vivo recordings
Gustatory receptor 8a	Gr8a	CG15371		L-canavanine	
Gustatory receptor 9a	Gr9a	CG32693			
Gustatory receptor 89a	Gr89a	CG14901		bitter	
Gustatory receptor 98a	Gr98a	CG13976			
Gustatory receptor 98d	Gr98d	CG31061			
Gustatory receptor 98c	Gr98c	CG31060			
Gustatory receptor 98b	Gr98b	CG31059			
Gustatory receptor 28a	Gr28a	CG13787			
Gustatory receptor 28b	Gr28b	CG13788	↑ Ca2+ (Xiang 2010)		larval reporter
Gustatory receptor 33a	Gr33a	CG17213		pheromone, DEET	
Gustatory receptor 66a	Gr66a	CG7189		pheromone, DEET, L-canavanine, caffeine	
Gustatory receptor 43a	Gr43a	CG1712	↑ Ca2+ (Miyamoto 2012)	fructose	In vivo calcium imaging
Gustatory receptor 59e	Gr59e	CG33151			
Gustatory receptor 59f	Gr59f	CG33150			
Gustatory receptor 57a	Gr57a	CG13441			
Gustatory receptor 47b	Gr47b	CG30030			
Gustatory receptor 68a	Gr68a	CG7303		mating pheromone	
Gustatory receptor 32a	Gr32a	CG14916		mating pheromone, bitter	
Gustatory receptor 2a	Gr2a	CG18531			
Gustatory receptor 39a	Gr39a	CG31622		mating pheromone, bitter	
Gustatory receptor 39b	Gr39b	CG31620			
Gustatory receptor 23a	Gr23a	CG15396			

Appendix B:

Supplementary Materials from Chapter 5

Figure S1: EMSAs of additional CrebA binding sites. For each SPCG, the enhancer region is presented with orange circles representing predicted CrebA binding sites. Green circles indicate gel shifts shown in Fig. 1. In each case, site 1 is to the left. Lanes are numbered from 1-10, or 1-11, and the key is listed at the bottom of the page. CrebA binds to each site and is competed away by unlabeled oligonucleotides that correspond to the same sequence, in a majority of examples. When the binding sequence is mutated (red) in the unlabled oligos, they are no longer able to compete for CrebA binding.

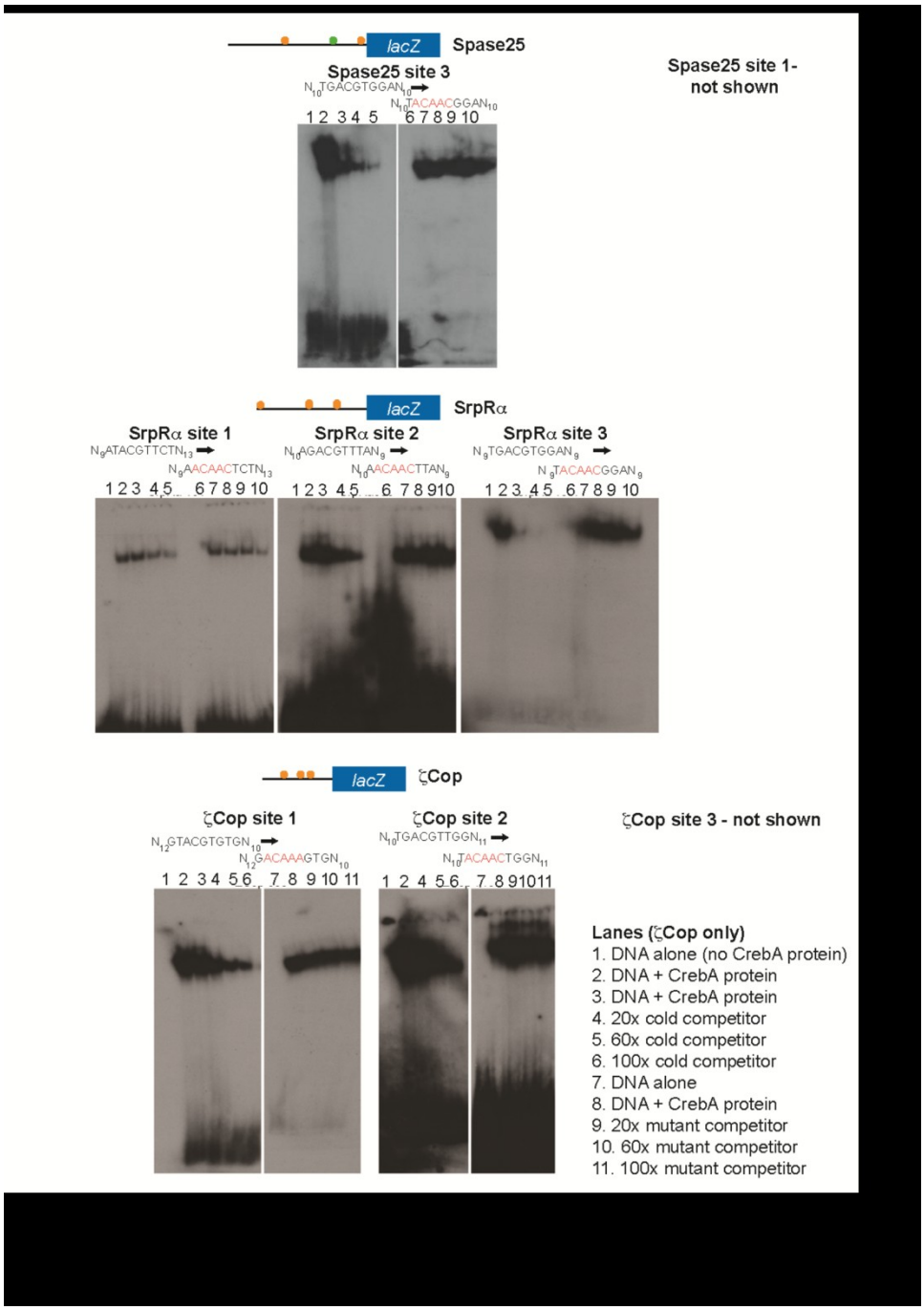
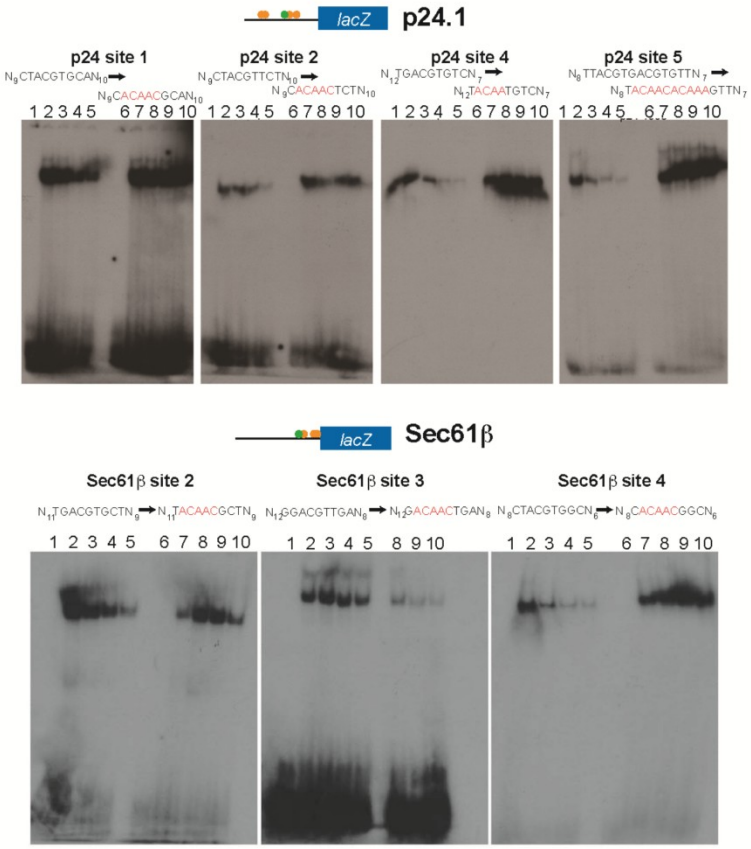

lacZ
Spase25
Spase25 site 3
N10TGACGTGGAN10
N10TACAACGGAN10
1 2 3 4 5 6 7 8 9 10
Spase25 site 1- not shown
lacZ
SrpRα
SrpRα site 1
N9ATACGTTCTN13
N9AACAACTCTN13
1 2 3 4 5 6 7 8 9 10
SrpRα site 2
N10AGACGTTTAN9
N10AACAACTTAN9
1 2 3 4 5 6 7 8 9 10
SrpRα site 3
N9TGACGTGGAN9
N9TACAACGGAN9
1 2 3 4 5 6 7 8 9 10
lacZ
ζCop
ζCop site 1
N12GTACGTGTGN10
N12GACAAAGTGN10
1 2 3 4 5 6 7 8 9 10 11
ζCop site 2
N10TGACGTTGGN11
N10TACAACTGGN11
1 2 4 5 6 7 8 9 10 11
ζCop site 3 - not shown
Lanes (ζCop only)
1. DNA alone (no CrebA protein)
2. DNA + CrebA protein
3. DNA + CrebA protein
4. 20x cold competitor
5. 60x cold competitor
6. 100x cold competitor
7. DNA alone
8. DNA + CrebA protein
9. 20x mutant competitor
10. 60x mutant competitor
11. 100x mutant competitor

Figure S1



Lanes

1. DNA alone (no CrebA protein)
2. DNA + CrebA protein
3. 20x cold competitor
4. 60x cold competitor
5. 100x cold competitor
6. DNA alone
7. DNA + CrebA protein
8. 20x mutant competitor
9. 60x mutant competitor
10. 100x mutant competitor

Supplemental Table 1: List of genes regulated by CrebA and associated human orthologues

Table S1		
CrebA targets - GO	**Human Homolog**	**Fold change**
Ion Channels/transporters/binding proteins		
CG34123 - TRPA1-like channel/thermotaxis	**TRPM3**	**-1.50**
CG5427 (Oatp33Ea) - Organic anion transporter		**-1.54**
CG9467 - voltage gated potassium channel	**KCTD3**	**-1.59**
CG6356 - organic cation transporter		**-1.62**
CG7571 (Oatp74D) - Organic anion transporter		**-1.66**
CG10449 (catsup) - metal ion transmembrane transporter activity	**SLC39A7 and SLC39A4**	**-1.76**
CG10620 (Tsf2) - iron ion transmembrane transporter activity	**MFI2**	**-1.82**
CG9261 (nrv2) - sodium:potassium-exchanging ATPase activity		**-1.90**
CG6672 - zinc ion transmembrane transporter activity	**SLC30A7**	**-2.07**
CG2177 - metal ion transmembrane transporter activity	**SLC39A9**	**-2.23**
CG15094 - high affinity inorganic phosphate:sodium symporter activity		**-3.28**
CG18039 (KaiRIA) - ionotropic glutamate receptor activity		**-3.44**
CG32373 - EGF-like Ca+2 ion binding		**-1.51**
CG30104- hydrolase activity, acting on ester bonds		**-1.67**
CG17271 - calcium ion binding		**-2.55**
Transporters		
CG31103 - sugar transporter		**-1.57**
CG14621 - glucose 6-phosphate:phosphate antiporter	**SLC35E1**	**-1.65**
CG9706 - acetyl-CoA transporter activity	**SLC33A1**	**-1.68**
CG4484 - sucrose:hydrogen symporter activity	**SLC45A1**	**-1.69**
CG13795 - neurotransmitter transporter activity		**-1.93**
CG15438 - sodium-dependent phosphate transmembrane transporter activity	**SLC17A9**	**-2.10**
CG8249 - glucose transmembrane transporter activity		**-2.14**
CG3424 (path) - amino acid transmembrane transporter activity		**-3.02**
CG2675 (Csat) - UDP-galactose transmembrane transporter activity	**SLC35A2**	**-3.53**
CG8100 - oxygen transporter activity		**-6.53**
Cell death/apoptosis		
CG18593 (viaf) - regulation of caspase activity	**PDCL3**	**-1.54**
CG32491 (mod(mdg4)) - BTB domain, chromatin assembly, regulation of apoptosis		**-1.56**
CG12297 (BG4) - death receptor binding		**-1.79**
CG5123 (W) - anatomical structure development; programmed cell death		**-2.00**
CG1274 (Jafrac2) - peroxidase activity, induction of apoptosis	**PRDX4**	**-2.47**

Nucleic acid binding (DNA and RNA)		
CG11033 - similar to vertebrate jumonji, zinc ion binding		**-1.51**
CG11494 (BtbVII) - BTB domain		**-1.51**
CG10053 - nucleic acid binding	**CCDC75**	**-1.51**
CG8631 (msl-3) - chromatin/mRNA binding, transcriptional regulator	**MSL3**	**-1.85**
CG8205 (fus) - mRNA/nuclei acid binding, EGFR signaling	**ESRP1**	**-1.52**
CG5147 - DNA binding, RNA polymerase III	**POLR3D**	**-1.56**
CG12863 - DNA binding, zinc-finger domain	**ZCCHC4**	**-1.58**
CG8730 (Drosha) - RNA polymerase III, miRNA processing	**RNASEN**	**-1.60**
CG15552 (Sox100B) - DNA bending, transcription factor activity	**SOX8**	**-1.60**
CG8676 (Hr39) - hormone receptor, transcription factor		**-1.61**
CG3019 (su(w[a])) - mRNA splicing, RNA processing	**SFRS8**	**-1.61**
CG12952 (sage) - transcription factor	**ATOH8**	**-1.96**
CG31365 - DNA binding, zinc finger domain		**-1.63**
CG4360 - dna binding, zinc finger domain		**-1.64**
CG2244 (MTA1-like) - DNA binding, Zinc finger, transcription factor	**MTA1**	**-1.64**
CG8920 - contains Tudor domain often found in RNA-binding proteins	**TDRD7**	**-1.94**
CG6203 (Fmr1) - mRNA binding	**FXR2 and FXR1**	**-1.70**
CG6634 (mid) - transcription factor		**-1.72**
CG5460 (H) - transcriptional co-repressor		**-1.72**
CG1650 (unpg) - transcription factor activity	**GBX2**	**-1.74**
CG1070 (Alh) - transcription factor activity	**MLLT10**	**-1.80**
CG12245 (gcm) - DNA binding; transcription factor activity	**GCM1**	**-1.86**
CG3886 (Psc) - DNA binding		**-1.86**
CG30291 (grau) - specific RNA polymerase II transcription factor activity		**-1.87**
CG1856 (ttk) - specific RNA polymerase II transcription factor activity		**-1.92**
CG7008 (Tudor-SN) - transcription coactivator activity	**SND1**	**-1.93**
CG11680 (mle) - ATP-dependent helicase activity	**DHX9**	**-2.05**
CG4717 (kni) - ligand-dependent nuclear receptor activity; transcription factor activity		**-2.18**
CG8089 - zinc ion binding; nucleic acid binding		**-2.20**
CG12296 (klu) - nucleic acid binding; zinc ion binding		**-2.51**
CG1414 (bbx) - DNA binding; transcription factor activity	**BBX**	**-2.54**
CG11205 (phr) - deoxyribodipyrimidine photo-lyase activity		**-2.75**
CG10352 (Rat1) - 5'-3' exoribonuclease activity		**-2.81**
CG17077 (pnt) - transcription factor activity	**ETS1**	**-4.50**
CG9989 - nucleic acid binding; metal ion binding		**-5.21**
CG7450 (CrebA) - regulation of transcription	**CREB3L1 and CREB3L2**	**-13.65**

RNA processing		
CG8038 - RNA binding, rRNA/tRNA processing, mRNA cleavage	**POP4**	**-1.51**
CG9735 (Aats-trp) - tryptophan-tRNA ligase	**WARS**	**-1.62**
CG32281 - tRNA methyltransferase activity	**TRMT5**	**-1.95**
CG5394 (Aats-glupro) - glutamate-tRNA ligase activity	**EPRS**	**-1.78**
Unknown		
CG6559		**-1.51**
CG8230 - contains Dymeclin domain	**DYM (Dymeclin)**	**-1.52**
CG13936		**-1.52**
CG33052	**(Rat) GORAB_RAT**	**-1.53**
CG9669	**C11orf10**	**-1.55**
CG11314 - mesoderm development, similar to Neimann Pick Type C protein		**-1.55**
CG8997		**-1.55**
CG8780 (tey) - involved in muscle development		**-1.56**
CG7161 (Oseg1) - sensory cilium assembly	**IFT122**	**-1.58**
CG34331		**-1.59**
CG42336		**-1.60**
CG11395		**-1.60**
CG42259		**-1.61**
CG5174	**TPD52L2**	**-2.11**
CG5850	**TMEM184C**	**-1.61**
CG14905	**CCDC63**	**-1.62**
CG34276		**-1.62**
CG6325	**DCAF15**	**-1.62**
CG31606		**-1.63**
CG14646	**TMEM129**	**-1.64**
CG14629		**-1.65**
CG5768		**-2.38**
CG13857		**-1.66**
CG13565		**-1.66**
CG9257	**MLEC (Malectin)**	**-1.68**
CG7510 - contains Plekstrin domain, cell signaling	**GPR155**	**-1.68**
CG12063 - likely involved in carbohydrate binding		**-1.68**
CG5021	**FAM18A**	**-1.68**
CG10616	**C1orf27**	**-1.69**
CG33169		**-1.84**
CG32243		**-1.72**
CG3767 (Jhl-26)		**-1.73**
CG6785		**-1.74**
CG6983		**-1.75**
CG13614		**-1.79**
CG9196 (Spz6) - related to Spatzle		**-1.81**
CG5862	**DDRGK1**	**-1.81**
CG8613 - catalytic activity	**SPATA20**	**-1.81**
CG13082		**-1.82**
CG1104	**KIAA0776**	**-1.82**
CG11018		**-1.83**
CG14454		**-2.53**
CG8717 (slv) - likely transmembrane	**RAG1AP1**	**-1.83**

CG8331	**REEP5**	**-1.84**
CG10591		**-1.84**
CG5484	**YIF1B**	**-1.86**
CG9945	**WDR23**	**-1.87**
CG14265		**-1.87**
CG7802		**-1.89**
CG14400		**-1.89**
CG33003		**-1.89**
CG10157		**-1.89**
CG2812		**-1.90**
CG13394		**-1.93**
CG14110		**-1.93**
CG1077		**-1.95**
CG15905		**-1.99**
CG11905		**-2.67**
CG33272 - is located in the glue gene cluster at 68C		**-2.02**
CG12384		**-2.09**
CG5039		**-2.09**
CG4844		**-2.12**
CG12011		**-2.12**
CG14720		**-2.13**
CG7300		**-2.15**
CG9186	**C2orf43**	**-2.17**
CG7011	**ERGIC3**	**-2.18**
CG34224		**-2.18**
CG4293	**ERGIC2**	**-2.18**
CG13362		**-2.20**
CG13822		**-2.23**
CG15024		**-2.32**
CG7137	**RRP8**	**-2.32**
CG8145		**-2.33**
CG34035		**-2.36**
CG18661	**C16orf13**	**-2.43**
CG30272		**-2.50**
CG5476 (TwdlN)		**-2.51**
CG6478 (TwdlB)		**-2.77**
CG6447 (TwdlL)		**-2.77**
CG7567		**-2.99**
CG32039	**SVIP**	**-3.13**
CG14756		**-3.23**
CG18343		**-3.73**
CG10717 (ImpL1) - Ecdysone-inducible factor 1		**-4.14**
CG32368		**-5.46**
CG7936 (mex1) - midgut expression		**-14.33**
Other		
CG9704 (Nrt) - axon guidance, axonogenesis		**-1.76**
CG32282 (dro4) - defense response to fungus		**-1.51**
Intracellular proteolysis		
CG6726 - metallopeptidase		**-1.59**
CG10576 - aminopeptidase	**PA2G4**	**-1.65**

CG7649 (Neu3) - metalloendopeptidase activity		**-1.81**
CG42335 - aminopeptidase activity; metallopeptidase activity		**-4.29**
CG6372 - aminopeptidase activity; metalloexopeptidase activity		**-6.93**
Extracellular proteolysis		
CG3344 - serine type carboxypeptidase	**SCPEP1**	**-1.51**
CG10992 - cysteine endopeptidase	**CTSB**	**-1.57**
CG13744 - serine endopeptidase		**-1.60**
CG17633 - metallocarboxypeptidase activity		**-1.68**
CG8539 - metallocarboxypeptidase activity; zinc ion binding		**-1.83**
CG9460 - serine-type endopeptidase inhibitor		**-1.88**
CG32483 - serine-type carboxypeptidase activity		**-2.20**
CG8579 (Jon44E) - serine-type endopeptidase activity		**-4.65**
Cell cycle		
CG8598 (eco) - N-acetyltransferase	**ESCO2**	**-1.51**
CG10800 (Rca1) - regulation of mitosis		**-1.52**
CG1395 (Stg) - protein tyrosine phosphatase	**CDC25B/C**	**-1.69**
CG8857 (RpS11) - ribosomal subunit	**RPS11P5**	**-1.71**
CG32417 (Myt1) - protein serine/threonine/tyrosine kinase activity	**PKMYT1**	**-1.77**
CG30291 - regulation of cyclin-dependent kinase activity	**CDK5RAP3**	**-1.87**
CG17064 (mars) - regulation of mitotic cell cycle		**-1.90**
CG12740 (RpL28) - structural constituent of ribosome	**RPL28**	**-3.58**
Protein Binding		
CG3408 - containe leucine-rich repeat	**LRRC59**	**-1.56**
CG31274 - contains leucine-rich repeat		**-1.59**
CG31447 (MESK4) - contains leucine-rich repeat		**-1.59**
CG14351 - contains leucine-rich repeat		**-1.58**
CG5692 (raps) - GTPase activator involved in cell division	**GPSM2**	**-1.65**
CG32356 (ImpE1) - contains LDLR domain		**-2.14**
CG30023 (Sprt) - contains PDZ domain		**-1.69**
CG4195 (l(3)73Ah) - zinc ion binding	**PCGF3**	**-1.71**
CG2198 (Ama) - cell adhesion, antigen binding	**LSAMP and NEGR1**	**-1.76**
CG13125 - protein phosphatase type 1 regulator activity	**LRRC48**	**-1.77**
CG10420 - contains Armadillo-helical domain	**SIL1**	**-1.87**
CG11714 - contains BTB domain		**-1.95**
CG1418 - Rab GTPase binding	**RABAC1**	**-2.51**
CG1410 (waw) - GTPase activity	**GUF1**	**-2.54**
CG30483 (Prosap) - protein binding	**SHANK1**	**-2.90**
CG16757 (Spn) - protein phosphatase 1 binding	**PPP1R9B**	**-3.01**
CG33983 (obst-H) - chitin binding		**-4.93**
Cytoskeleton		
CG11242 - contains cytoskeleton associated CAP-Gly domain	**TBCB**	**-1.52**

CG8261 (Ggamma1) - actin organization	**GNG12**	**-1.52**
CG13913 (mwh) - actin binding		**-1.73**
CG32149 (RhoGAP71E) - signal transduction		**-2.16**
Ubiquitin/proteasome		
CG11887 (StlP) - proteasome assembly/activator	**ELP2**	**-1.52**
Mitochondria/Peroxisome		
CG10622 (Sucb) - succinate CoA ligase, tricarboxylic acid cycle	**SUCLG2**	**-1.53**
CG32174 - contains Coenzyme Q 4 domain	**COQ4**	**-1.53**
CG6439 - isocitrate dehydrogenase (NAD+) activity	**IDH3B**	**-1.65**
CG9961 - phosphoglycerate kinase activity, glycolysis		**-1.78**
CG5904 (mRpS31) - mitochondrial ribosomal protein	**MRPS31**	**-1.84**
CG13827 - peroxisome fission	**PEX11G**	**-1.84**
Receptors/Signal transduction		
CG18859 (Or19a) - olfactory receptor, odorant binding		**-1.56**
CG32825 (Or19b) - olfactory receptor, odorant binding, G-protein signaling		**-1.56**
CG4626 (Fz4) - frizzled 4 receptor		**-1.58**
CG8967 (Otk) - cell adhesion molecule binding	**PTK7**	**-1.71**
CG17592 (LBR) - lamin B receptor	**USF2**	**-1.85**
CG5912 (arr) - Wnt receptor signaling pathway	**LRP6**	**-1.90**
CG6127 (Ser) - epidermal growth factor receptor binding	**JAG2**	**-2.03**
CG34449 - receptor activity	**ZDHHC8**	**-2.58**
CG6965 (mthl5) - G-protein coupled receptor activity		**-2.15**
CG32475 (mthl8) - G-protein coupled receptor protein signaling pathway		**-2.84**
CG6104 (m2) - Notch signaling pathway		**-1.68**
CG8337 (malpha) - Notch signaling pathway		**-1.91**
Metabolism		
CG11899 - O-phospho-L-serine:2-oxoglutarate aminotransferase	**PSAT1**	**-1.59**
CG6673 - glutathione transferase activity		**-1.60**
CG5171 - trehalose phosphatase activity		**-1.60**
CG17560 - catalytic activity, male sterility		**-1.61**
CG6870 - electron carrier, heme binding		**-1.63**
CG18559 (Cyp309a2) - electron carrier, heme binding, monooxygenase activity		**-1.63**
CG8036 - transketolase activity	**TKT**	**-1.65**
CG7685 - alpha glucosidase activity, contains LDLR domain		**-1.67**
CG7623 (sll) - UDP-galactose transmembrane transporter activity	**SLC35B2**	**-1.67**
CG3534 - xylulokinase activity/carbohydrate metabolism	**XYLB**	**-1.51**
CG17026 - Inositol-monophosphate activity		**-1.72**
CG17562 - catalytic activity, male sterility		**-1.72**
CG6578 (Phm) - ecdysteroid 25-hydroxylase activity		**-1.73**

CG10621 - selenocysteine methyltransferase activity		**-1.73**
CG1041 - carnitine O-acetyltransferase activity	**CRAT**	**-1.73**
CG10361 - glycine C-acetyltransferase activity	**GCAT**	**-1.74**
CG9523 - protein adenylyltransferase activity	**FICD**	**-1.78**
CG14670 (Hcs) - biotin-[acetyl-CoA-carboxylase] ligase activity	**HLCS**	**-1.81**
CG8652 (Ugt37c1) - glucuronosyltransferase activity		**-1.86**
CG10924 - phosphoenolpyruvate carboxykinase (GTP) activity		**-1.88**
CG11052 - acylphosphatase activity		**-1.89**
CG1969 - glucosamine 6-phosphate N-acetyltransferase activity	**GNPNAT1**	**-2.04**
CG33503 (Cyp12d1-d) - electron carrier activity		**-1.95**
CG5946 - cytochrome-b5 reductase activity	**CYB5R3**	**-1.95**
CG6128 (Fuca) - alpha-L-fucosidase activity	**FUCA2**	**-1.96**
CG10512 - oxidoreductase activity		**-1.97**
CG10467 - aldose 1-epimerase activity; carbohydrate binding	**GALM**	**-2.00**
CG33093 - oxidoreductase activity		**-2.00**
CG10391 (Cyp310a1) - electron carrier activity; heme binding		**-2.12**
CG2062 (Cyp4e1) - electron carrier activity		**-2.15**
CG5656 - alkaline phosphatase activity	**ALPL**	**-2.17**
CG18585 - metallocarboxypeptidase activity	**CPA1 and CPA2**	**-2.27**
CG4899 (Pdh) - alcohol dehydrogenase (NAD) activity	**HPGD**	
CG10116 - lipoprotein lipase activity, lipid metabolic process		**-2.48**
CG9449 - acid phosphatase, phagocytosis, engulfment		**-2.81**
CG16799 - lysozyme activity, defense response; antimicrobial humoral response		**-2.65**
CG15362 - catalytic activity		**-2.76**
CG31809 - steroid dehydrogenase activity		**-2.95**
CG12242 (GstD5) - glutathione -S- transferase		**-3.07**
CG15743 - 3'(2'),5'-bisphosphate nucleotidase activity	**IMPAD1**	**-3.14**
CG1749 - Mo-molybdopterin cofactor sulfurase activity	**UBA5**	**-3.44**
CG31313 - cysteine-type endopeptidase inhibitor activity		**-4.00**
CG18088 - catalytic activity, metabolic process		**-6.72**
CG2759 (w) - biogenic amine biosynthetic process		**-3.80**
Secretory pathway		
ER-Golgi trafficking/transport		
CG8605 - RINT1/TIP1 domain	**RINT1**	**-1.51**
CG7966 - may function as selenium binding protein which is thought to function in the late stages of Golgi transport	**SELENBP1**	**-1.54**
CG32654 (Sec16) - protein exit from ER	**SEC16A**	**-1.54**
CG12404 - Yip1 domain	**YIPF5**	**-1.56**
CG4645 - Yip1 domain	**YIPF1**	**-1.60**
CG4848 - Vps51/Vps67 domin	**COG1**	**-1.60**
CG3652 - Yip1 domain	**YIPF6**	**-1.64**
CG5183 (KdelR) - KDEL receptor	**KDELR1**	**-1.68**

CG31040 (Cog7) - conserved oligomeric Golgi complex	**COG7**	**-1.71**
CG1528 (gammaCop) - retrograde vesicle-mediated transport	**COPG**	**-1.74**
CG6223 (betaCop) - retrograde vesicle-mediated transport	**COPB1**	**-1.89**
CG6699 (beta'Cop) - ER to Golgi vesicle-mediated transport	**COPB2**	**-1.90**
CG7961 (alphaCop) - retrograde vesicle-mediated transport	**COPA**	**-2.11**
CG2023 - Sec20 domain	**BNIP1**	**-2.31**
Endocytosis/membrane formation		
CG1049 (Cct) - choline phosphate cytidyltransferase activity	**PCYT1A**	**-1.53**
Golgi modification enzymes/proteins		
CG3874 (frc) - sugar transmembrane transporter	**SLC35D1**	**-1.53**
CG12030 - UDP-glucose-4 epimerase activity	**GALE**	**-1.57**
CG32775 (GlcAT-I) - glucuronosyltransferase activity	**B3GAT3**	
CG31002 - glucuronosyltransferase activity		**-1.64**
CG9614 (pip) - heparan sulfate 2-O-sulfotransferase activity		**-4.16**
CG12311 (tw) - dolichyl-phosphate-mannose-protein mannosyltransferase activity	**POMT2**	**-1.76**
CG10166 - dolichyl-phosphate beta-D-mannosyltransferase activity, protein amino acid glycosylation	**DPM1**	**-1.77**
CG6822 (ergic53) - mannose binding		**-2.46**
CG2103 (pgant6) - polypeptide N-acetylgalactosaminyltransferase	**GALNT10**	**-2.09**
Possible cargo		
CG8050 (Cys) - Cystatin, protease inhibitor		**-1.53**
CG5634 (dsd) - contains EGF-like 3 domain found in proteins known to be secreted		**-1.60**
CG13744 - serine endopeptidase - these enzymes are secreted		**-1.60**
CG32400 (Lcp65Ab1) - structural constituent of chitin-based larval cuticle		**-1.66**
CG32400 (Lcp65Ab2) - structural constituent of chitin-based larval cuticle		**-1.66**
CG1567 (C901) - contains EGF-like 3 domain	**DLK2**	**-1.67**
CG2560 (Cpr11A) - structural constituent of chitin-based larval cuticle		**-1.69**
CG14089 - contains collagen triple helix repeat		**-1.71**
CG4914 - serine endopeptidase - these enzymes are secreted		**-1.71**
CG5276 - member of Apyrase family of proteins		**-1.76**
CG2044 (Lcp4) - structural constituent of chitin-based larval cuticle		**-1.83**
CG9307 (Cht5) - chitinase activity		**-1.85**
CG9355 (dy) - structural constituent of chitin-based cuticle		**-2.01**

CG17044 (yellow-e2) - member of Major Royal Jelly family.		**-2.10**
CG2043 (Lcp3) - structural constituent of chitin-based larval cuticle		**-2.37**
CG18348 (Cpr67Fb) - structural constituent of chitin-based cuticle		**-2.38**
CG9120 (LysX) - lysozyme activity		**-2.65**
CG7941 (Cpr67Fa1) - structural constituent of chitin-based larval cuticle		**-2.98**
CG34270 (Cpr65Ax1) - structural constituent of chitin-based cuticle		**-3.04**
CG18777 (Cpr65Ax2) - structural constituent of chitin-based cuticle		**-3.04**
CG7663 (Cpr78Cb) - structural constituent of chitin-based cuticle		**-3.38**
CG33117 (Victoria) - member of Drosophila species specific Turandot humoral factor		**-3.70**
CG9070 (Cpr47Eg) - structural constituent of chitin-based cuticle		**-4.43**
CG10533 (Lcp65Af) - structural constituent of chitin-based larval cuticle		**-4.44**
CG18349 (Cpr67Fa2) - structural constituent of chitin-based cuticle		**-1.82**
CG6912 - contains growth factor receptor domain, mucin like (runs of T)		**-2.16**
CG14850 - mucin like (30.2% TS)		**-2.20**
CG32453 - mucin-like (39.6% TS)		**-2.53**
CG12546 - mucin (short 43.4% TS)		**-2.58**
CG14452 - mucin (short 44.3% TS)		**-2.58**
CG32073 - mucin (short 41.2% TS)		**-3.19**
CG13738 - mucin (short 31.1% TS)		**-4.28**
CG5402 - mucin (short 31.4% TS)		**-4.37**
CG14453 - mucin (short 40.9% TS)		**-4.58**
ER proteins		
CG9726 (PH4alphaMP) - procollagen proline-4 dioxygenase activity		**-1.54**
CG15818 - contains C-type lectin domain		**-1.52**
CG31014 (PH4alphaSG1) - procollagen proline-4 dioxygenase activity		**-1.54**
CG16905 (eloF) - fatty acid elongase of the GNS1/SUR4 family		**-1.58**
CG5417 (Srp14) - signal recognition particle	**SRP14**	**-1.59**
CG4758 (Trp1) - SRP-dependent cotranslational protein targeting	**SEC62**	**-1.63**
CG7748 (OstStt3) - oligosaccharyl transferase activity	**STT3B**	**-1.72**
CG7556 - unfolded protein binding	**DNAJC1**	**-1.74**
CG9911 - protein disulfide isomerase activity	**ERP44**	**-1.75**
CG2469 - contains tetratricopeptide motif	**CTR9**	**-1.76**
CG9539 (Sec61alpha) - protein transporter activity	**SEC61A2**	**-1.78**
CG5434 (Srp72) - SRP-dependent cotranslational protein targeting	**SRP72**	**-1.81**
CG7872 - contains DnaJ domain involved in protein	**DNAJC25**	**-1.81**

folding		
CG9459 - contains GNS1/Sur4 domain		**-1.85**
CG5520 (Gp93) - unfolded protein binding	**HSP90B1**	**-1.87**
CG8860 - protein transporter activity	**SEC61G**	**-1.88**
CG2522 (Gtp-bp) - signal recognition particle binding	**SRPR**	**-1.88**
CG4457 (Srp19) - SRP-dependent cotranslational protein targeting	**SRP19**	**-1.89**
CG5474 (SsRbeta) - protein retention in ER lumen	**SSR2**	**-1.97**
CG10130 (Sec61beta) - SRP-dependent cotranslational protein targeting	**SEC61B**	**-1.99**
CG8268 (Srp9) - SRP-dependent cotranslational protein targeting	**SRP9 and SRP9L1**	**-1.99**
CG5885 - cotranslational protein targeting to membrane	**SSR3**	**-2.02**
CG8583 (sec63) - signal recognition particle binding	**SEC63**	**-2.03**
CG5064 (Srp68) - SRP-dependent cotranslational protein targeting	**SRP68**	**-2.04**
CG8420 - contains ER targeting motif	**GRLF1**	**-2.05**
CG9342 (Mtp) - phosphatidylcholine transmembrane transporter	**MTTP**	**-2.06**
CG32276 - protein modification process	**SERP1**	**-2.07**
CG14214 (Sec61gamma) - SRP-dependent cotranslational protein targeting		**-2.08**
CG9035 (Tapdelta) - protein retention in ER lumen	**SSR4**	**-2.12**
CG18431 - contains C-type lectin domain		**-2.12**
CG1751 (Spase25) peptidase activity, signal peptide processing	**SPCS2**	**-2.19**
CG9302 - protein disulfide isomerase activity	**PDIA5**	**-2.24**
CG2358 (Spase18-21) - signal peptide processing	**SEC11B and SEC11C**	**-2.30**
CG3810 (Edem1) - ER-associated protein catabolic process	**EDEM2**	**-2.42**
CG11840 (Spp) - membrane protein proteolysis	**HM13**	**-2.50**
CG11500 (Spase12) - signal peptide processing	**SPCS1**	**-2.55**
CG9698 - procollagen-proline 4-dioxygenase activity		**-2.83**
CG12918 - contains ER targeting motif	**CNPY2**	**-2.87**
CG14105 - contains tetratricopeptide motif	**TTC36**	**-3.25**
CG33162 (SrpRbeta) - SRP-dependent cotranslational protein targeting to membrane	**SRPRB**	**-3.31**
CG11642 (TRAM) - SRP-dependent cotranslational protein targeting	**TRAM1**	**-5.22**
CG7945 - unfolded protein binding	**BAG2**	**-7.05**
Vesicle/vesicle transport		
CG5359 - contains Tctex-1 domain	**TCTEX1D2**	**-1.54**
CG13626 (Syx18) - SNAP receptor	**STX18**	**-1.55**
CG5341 (Sec6) - exocytosis	**EXOC3**	**-1.65**
CG33523 - structural molecule activity	**MOSPD2**	**-1.78**
CG8266 (Sec31) - component of COP II vesicle	**SEC31A**	**-1.82**
CG33105 (p24-2) - vesicle transport		**-1.90**
CG3948 (zetaCop) - intracellular protein transport	**COPZ1**	**-2.69**
CG3564 (CHOp24) - transport	**TMED2**	**-2.21**
CG11785 (bai) - vesicle-mediated transport	**TMED10**	**-2.24**
CG33104 (eca) - transport	**TMED4**	**-2.25**

CG6773 (Sec13) - larval chitin-based cuticle development	**SEC13**	**-2.44**
CG10733 (loj) - transport, member of p24 family		**-2.51**
Golgi morphology		
CG6838 - ARF GTPase activator	**ARFGAP3**	**-1.58**
CG4237 (Gap69C) - ARF GTPase activator activity	**ARFGAP1**	**-1.95**
CG7809 (Grasp65) - Golgi organization	**GORASP2**	**-2.38**
CG11061 (GM130) - Golgi organization	**GOLGA2**	**-2.40**

Supplemental Table 2: Clustering analysis of gene ontology terms for CrebA target genes

Gene Ontology term		Fold Enrichment	P-value
Annotation Cluster 1	**Enrichment Score: 15.12**		
Cotranslational protein targeting to membrane		31	3.1E-19
Protein targeting to membrane		28	3.0E-18
Protein targeting to ER		30.5	9.6E-18
SRP-dependent cotranslational protein targeting to membrane		30.5	9.6E-18
Protein targeting		3.9	3.1E-6
Annotation Cluster 2	**Enrichment Score: 10.63**		
Endoplasmic reticulum membrane		11.5	1.6E-11
Nuclear envelope – endoplasmic reticulum network		11.2	2.1E-11
Endoplasmic reticulum part		9.7	4.0E-11
Annotation Cluster 3	**Enrichment Score: 8.17**		
Secretory pathway		4.4	4.8E-9
Secretion by cell		4.3	5.2E-9
secretion		4.1	1.2E-8
Annotation Cluster 4	**Enrichment Score: 6.35**		
Protein localization		3	3.5E-9
Protein transport		3.1	3.6E-8
Establishment of protein localization		3	4.7E-8
Macromolecule localization		2.7	6.9E-8
Intracellular protein transport		3.2	7.2E-8
Cellular localization		2.1	2.0E-5
Intracellular transport		2.2	3.9E-5
Establishment of cellular localization		2.1	7.4E-5
Annotation Cluster 5	**Enrichment Score: 4.23**		
Intracellular protein transport across a membrane		32.7	5.8E-6
SRP-dependent cotranslational protein targeting to membrane, translocation		32.7	5.8E-6
translocation		10.7	5.9E-3
Annotation Cluster 6	**Enrichment Score: 3.78**		
COPI coated vesicle membrane		25	1.4E-6
COPI vesicle coat		25	1.4E-6
COPI-coated vesicle		22.5	2.8E-6
Vesicle coat		8.3	6.3E-4
Coated vesicle membrane		8	7.5E-4
Cytoplasmic vesicle membrane		8	7.5E-4
Coated membrane		7.5	1.0E-3
Membrane coat		7.5	1.0E-3
Cytoplasmic vesicle part		6.4	2.1E-3
Vesicle membrane		5.9	3.1E-3
Annotation Cluster 7	**Enrichment Score: 3.58**		
Structural constituent of chitin-based cuticle		4.4	8.4E-5
Structural constituent of cuticle		3.7	3.9E-4
Insect cuticle protein		4.2	5.6E-4
Annotation Cluster 8	**Enrichment Score: 3.14**		
Golgi-associated vesicle		10.9	9.9E-7
Cytoplasmic membrane-bound vesicle		3.4	2.3E-3

Cytoplasmic vesicle	3.4	2.5E-3
Coated vesicle	3.7	2.6E-3
Membrane-bound vesicle	3.3	3.0E-3
Vesicle	3.3	3.2E-3

Supplemental Table 3. List of genes upregulated by Creb3L1 T expression in HeLa cells.

Table S3					
Transcript ID	gene_assignment	Gene Name	Description	p-value(Creb3L1 OE vs. Wild type)	Linear FC (Creb3L1 OE vs. Wild type)
8068713	NM_002462	MX1	myxovirus (influenza virus) resistance 1, interferon-inducib	0.0021922	17.0754
8068697	NM_002463	MX2	myxovirus (influenza virus) resistance 2 (mouse)	0.0111784	15.1
7939642	NM_052854	CREB3L1	cAMP responsive element binding protein 3-like 1	3.54E-05	12.9329
7958913	NM_002535	OAS2	2'-5'-oligoadenylate synthetase 2, 69	0.0103203	11.7111
8074606	NM_017414	USP18	ubiquitin specific peptidase 18	0.00223429	11.0737
7976443	NM_001130080	IFI27	interferon, alpha-inducible protein 27	0.0117214	9.18069
7914127	NM_002038	IFI6	interferon, alpha-inducible protein 6	0.000317565	6.69729
8071155	NM_017414	USP18	ubiquitin specific peptidase 18	0.0030251	6.31418
8090018	NM_031458	PARP9	poly (ADP-ribose) polymerase family, member 9	0.00107766	6.25609
8140971	NM_152703	SAMD9L	sterile alpha motif domain containing 9-like	0.015793	6.08203
7958895	NM_006187	OAS3	2'-5'-oligoadenylate synthetase 3, 100kDa	0.0057739	6.07502
7958884	NM_016816	OAS1	2',5'-oligoadenylate synthetase 1, 40	0.00545183	5.73649
7902541	NM_006820	IFI44L	interferon-induced protein 44-like	0.000170085	5.63117
8004184	NM_017523	XAF1	XIAP associated factor 1	0.0377372	5.15791
8120967	NM_002526	NT5E	5'-nucleotidase, ecto (CD73)	0.000553219	4.94924
8160559	NM_014314	DDX58	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58	0.00263519	4.89132
8056285	NM_022168	IFIH1	interferon induced with helicase C domain 1	0.0121087	4.86467
8113709	NM_002317	LOX	lysyl oxidase	0.00222634	4.58754
8140967	NM_017654	SAMD9	sterile alpha motif domain containing 9	0.00428013	4.48763
7929052	NM_001031683	IFIT3	interferon-induced protein with tetratricopeptide repea	0.00637832	4.44924
7929065	NM_001548	IFIT1	interferon-induced protein with tetratricopeptide repeats	0.01259	4.14387
8059650	NM_080424	SP110	SP110 nuclear body protein	0.0129056	4.11555
7916609	NM_002228	JUN	jun oncogene	0.00360672	4.03519
8103601	NM_001012967	DDX60L	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60-like	0.00221228	3.97904
7902290	NM_001902	CTH	cystathionase (cystathionine gamma-lyase)	0.000103572	3.93126
8103563	NM_017631	DDX60	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60	0.000178152	3.83915
8072735	NM_145343	APOL1	apolipoprotein L, 1	0.00150687	3.82463
8057744	NM_007315	STAT1	signal transducer and activator of transcription 1, 91kDa	0.00172396	3.76086

8007446	NM_005533	IFI35	interferon-induced protein 35	0.0262794	3.67129
7902553	NM_006417	IFI44	interferon-induced protein 44	5.76E-05	3.66502
8096335	NM_017912	HERC6	hect domain and RLD 6	0.000811332	3.66081
8043474	ENST00000390249	LOC652493	similar to Ig kappa chain V-I region HK102 precu	0.0323407	3.61282
8073015	NM_006855	KDELR3	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein rete	0.00115626	3.56966
8104901	NM_002185	IL7R	interleukin 7 receptor	0.000534183	3.52273
7937335	NM_003641	IFITM1	interferon induced transmembrane protein 1 (9-27)	0.0238214	3.5221
8082100	NM_017554	PARP14	poly (ADP-ribose) polymerase family, member 14	0.00608407	3.5212
8017850	NM_017983	WIPI1	WD repeat domain, phosphoinositide interacting 1	7.28E-05	3.50756
8086125	NM_014831	LBA1	lupus brain antigen 1	0.0231669	3.47905
7929047	NM_001547	IFIT2	interferon-induced protein with tetratricopeptide repeats	0.0140244	3.4529
7967117	NM_003733	OASL	2'-5'-oligoadenylate synthetase-like	0.0130991	3.39602
8087485	NM_003335	UBA7	ubiquitin-like modifier activating enzyme 7	0.0149622	3.26668
8082075	NM_138287	DTX3L	deltex 3-like (Drosophila)	0.00832646	3.12955
8067680	NM_001037335	PRIC285	peroxisomal proliferator-activated receptor A interac	0.0129138	3.1278
8051501	NM_002759	EIF2AK2	eukaryotic translation initiation factor 2-alpha kinase	0.00124567	3.125
8096361	NM_016323	HERC5	hect domain and RLD 5	0.000284829	3.12368
8131803	NM_000600	IL6	interleukin 6 (interferon, beta 2)	0.015383	3.07564
7917516	NM_002053	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa	0.0232525	3.0598
8091327	NM_021105	PLSCR1	phospholipid scramblase 1	0.0135583	3.04726
7919749	---			0.0179775	3.03499
8030383	NM_020650	RCN3	reticulocalbin 3, EF-hand calcium binding domain	0.00013582	3.03421
8063382	NM_005985	SNAI1	snail homolog 1 (Drosophila)	0.000118305	2.97849
8015511	NM_024119	DHX58	DEXH (Asp-Glu-X-His) box polypeptide 58	0.0147183	2.9618
7928944	NM_004670	PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	2.77E-05	2.94972
7973618	NM_006084	IRF9	interferon regulatory factor 9	0.000889813	2.91503
8143327	NM_022750	PARP12	poly (ADP-ribose) polymerase family, member 12	0.00519873	2.8755
8066117	NM_015474	SAMHD1	SAM domain and HD domain 1	0.00035986	2.86566
8048940	NM_001080391	SP100	SP100 nuclear antigen	0.00849364	2.85113
8072710	NM_030641	APOL6	apolipoprotein L, 6	0.000646228	2.84218

8156688	NM_014290	TDRD7	tudor domain containing 7	0.00521519	2.84136
7965979	NM_001034173	ALDH1L2	aldehyde dehydrogenase 1 family, member L2	0.0030909	2.76854
7945962	NM_003141	TRIM21	tripartite motif-containing 21	0.0366201	2.76705
8098611	NM_003265	TLR3	toll-like receptor 3	0.00435449	2.75319
7929072	NM_012420	IFIT5	interferon-induced protein with tetratricopeptide repeats	0.00450975	2.70753
8081288	NM_018004	TMEM45A	transmembrane protein 45A	3.52E-05	2.67039
7917503	NM_018284	GBP3	guanylate binding protein 3	0.0218868	2.66745
8104930	NM_004172	SLC1A3	solute carrier family 1 (glial high affinity glutamate tr	0.0368704	2.64958
8006531	NM_144975	SLFN5	schlafen family member 5	0.00875998	2.59655
8014316	NM_002985	CCL5	chemokine (C-C motif) ligand 5	0.0253862	2.58198
8127841	NM_015599	PGM3	phosphoglucomutase 3	0.0103295	2.57077
7964119	NM_005419	STAT2	signal transducer and activator of transcription 2, 113kDa	0.00874807	2.56567
8144802	NM_006207	PDGFRL	platelet-derived growth factor receptor-like	0.000185507	2.56427
8178115	NM_001710	CFB	complement factor B	0.0150625	2.51515
7902687	NM_001554	CYR61	cysteine-rich, angiogenic inducer, 61	0.00108136	2.47401
8179351	NM_001710	CFB	complement factor B	0.000536334	2.3897
8118345	NM_001710	CFB	complement factor B	0.0174145	2.37817
8117321	NM_006355	TRIM38	tripartite motif-containing 38	0.00820652	2.35905
7899394	NM_001105556	C1orf38	chromosome 1 open reading frame 38	0.0225317	2.34205
8025551	NM_018381	C19orf66	chromosome 19 open reading frame 66	0.0199122	2.28275
8163002	NM_004235	KLF4	Kruppel-like factor 4 (gut)	0.00237199	2.27858
8143279	NM_020119	ZC3HAV1	zinc finger CCCH-type, antiviral 1	0.00149803	2.26732
8042503	NM_002357	MXD1	MAX dimerization protein 1	0.00301124	2.24663
8118571	NM_002800	PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (lar	0.0354238	2.22365
8178211	NM_002800	PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (lar	0.0354238	2.22365
8179495	NM_002800	PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (lar	0.0354238	2.22365
7906400	NM_005531	IFI16	interferon, gamma-inducible protein 16	0.0315959	2.21595
8066905	NM_021035	ZNFX1	zinc finger, NFX1-type containing 1	0.0041036	2.21569
8016847	NM_005082	TRIM25	tripartite motif-containing 25	0.0112147	2.21537

7975676	NM_025057	C14orf45	chromosome 14 open reading frame 45	0.00348777	2.21065
8101701	NM_152542	PPM1K	protein phosphatase 1K (PP2C domain containing)	0.00916232	2.19234
8034696	---			0.00349098	2.18162
8013581	NM_174887	IFT20	intraflagellar transport 20 homolog (Chlamydomonas)	3.48E-05	2.17943
8105935	AK302597	POM121L8P	POM121 membrane glycoprotein-like 8 (rat) pseudogene	0.0250282	2.16915
8132503	NM_004760	STK17A	serine	0.00296709	2.14559
8170921	NM_017514	PLXNA3	plexin A3	0.00186295	2.13838
8147112	NM_198584	CA13	carbonic anhydrase XIII	0.000649078	2.13394
8175811	NM_152274	FAM58A	family with sequence similarity 58, member A	0.00106909	2.13149
7989501	NM_001218	CA12	carbonic anhydrase XII	0.00416888	2.10246
7991640	NM_152334	TARSL2	threonyl-tRNA synthetase-like 2	0.00165681	2.08048
8165334	NM_004669	CLIC3	chloride intracellular channel 3	0.00148262	2.07254
7987772	NM_139265	EHD4	EH-domain containing 4	0.00249192	2.06938
8101699	---			0.0304479	2.04719
7940775	NM_004585	RARRES3	retinoic acid receptor responder (tazarotene induced) 3	0.0228007	2.02702
8162729	NM_014788	TRIM14	tripartite motif-containing 14	0.00191372	2.02505
8102482	NM_014822	SEC24D	SEC24 family, member D (S. cerevisiae)	0.000174023	2.02408
7914342	NM_004102	FABP3	fatty acid binding protein 3, muscle and heart (mammary-de	0.00494765	2.01185
8112476	ENST00000360795	LOC100133280	similar to Putative POM121-like protein 1-lik	0.0488553	2.01087
8105348	NM_001008397	GPX8	glutathione peroxidase 8 (putative)	0.000107755	1.99218
8117458	NM_001145009	BTN3A1	butyrophilin, subfamily 3, member A1	0.0348601	1.9819
8114010	NM_002198	IRF1	interferon regulatory factor 1	0.0490342	1.97477
8035304	NM_004335	BST2	bone marrow stromal cell antigen 2	0.016716	1.97173
8135069	NM_000602	SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen	0.0131059	1.96838
7920531	NM_001111	ADAR	adenosine deaminase, RNA-specific	0.00411715	1.96721
8109194	NM_000112	SLC26A2	solute carrier family 26 (sulfate transporter), member 2	0.0269684	1.9663
7995838	NM_005952	MT1X	metallothionein 1X	0.00145882	1.96453
8108370	NM_001964	EGR1	early growth response 1	0.0278533	1.96229
7969626	NM_180989	GPR180	G protein-coupled receptor 180	0.00515594	1.94441

8150276	NM_001102559	PPAPDC1B	phosphatidic acid phosphatase type 2 domain containi	0.0104405	1.94382
7981290	NM_004184	WARS	tryptophanyl-tRNA synthetase	0.0101209	1.94013
8154135	NM_004170	SLC1A1	solute carrier family 1 (neuronal	0.0390481	1.93429
7900365	NM_001136493	MFSD2	major facilitator superfamily domain containing 2	0.0234807	1.91661
8159142	NM_000093	COL5A1	collagen, type V, alpha 1	0.00639825	1.91542
7958346	NM_152261	C12orf23	chromosome 12 open reading frame 23	0.00664511	1.91477
7919168	NM_022359	PDE4DIP	phosphodiesterase 4D interacting protein	0.000452123	1.91369
7909610	NM_001040619	ATF3	activating transcription factor 3	0.00177126	1.91347
8056408	NM_004482	GALNT3	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylga	0.000636847	1.90286
7975932	NM_020431	TMEM63C	transmembrane protein 63C	0.00179614	1.8979
8104693	NM_178140	PDZD2	PDZ domain containing 2	0.0330028	1.89742
7999909	NM_016235	GPRC5B	G protein-coupled receptor, family C, group 5, member B	0.00594844	1.8911
7956426	NM_031479	INHBE	inhibin, beta E	0.00469058	1.88127
7945462	NM_004031	IRF7	interferon regulatory factor 7	0.0375506	1.8682
7982094	NR_003359	SNORD115-44	small nucleolar RNA, C	0.0251693	1.86405
8018975	NM_005567	LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein	0.0174428	1.86079
8019622	NM_145041	TMEM106A	transmembrane protein 106A	0.00221408	1.85973
7969129	NM_001040443	PHF11	PHD finger protein 11	0.00268231	1.85963
8180359	---			0.000270765	1.85408
7909661	NM_012424	RPS6KC1	ribosomal protein S6 kinase, 52kDa, polypeptide 1	0.00195466	1.85353
8001147	NM_198490	RAB43	RAB43, member RAS oncogene family	0.000572548	1.85065
8075720	NM_030882	APOL2	apolipoprotein L, 2	0.0118681	1.84565
7926875	NM_012342	BAMBI	BMP and activin membrane-bound inhibitor homolog (Xenopus	0.00184362	1.84473
8029914	NM_018485	GPR77	G protein-coupled receptor 77	0.00578763	1.84323
7983157	NM_024956	TMEM62	transmembrane protein 62	0.0461437	1.83416
8094101	NM_001014448	CPZ	carboxypeptidase Z	0.000151658	1.83116
8007483	NM_145041	TMEM106A	transmembrane protein 106A	0.0018775	1.83034

7999173	NM_145253	FAM100A	family with sequence similarity 100, member A	0.0129583	1.82553
7985777	NM_002201	ISG20	interferon stimulated exonuclease gene 20kDa	0.011981	1.82378
8072659	NR_024194	TOM1	target of myb1 (chicken)	0.00495348	1.82108
8065136	NM_001042576	RRBP1	ribosome binding protein 1 homolog 180kDa (dog)	0.000839294	1.81869
7922773	NM_000433	NCF2	neutrophil cytosolic factor 2	9.40E-05	1.81838
8162744	NM_003389	CORO2A	coronin, actin binding protein, 2A	0.00144357	1.81753
7955694	NM_002178	IGFBP6	insulin-like growth factor binding protein 6	0.00464103	1.81605
8052331	NM_033109	PNPT1	polyribonucleotide nucleotidyltransferase 1	0.0486448	1.8159
7982056	NR_003342	SNORD115-25	small nucleolar RNA, C	0.0160967	1.81531
8103043	NM_001029998	SLC10A7	solute carrier family 10 (sodium	0.0100227	1.81062
8042283	NM_014181	HSPC159	galectin-related protein	0.00204248	1.80967
8041206	NM_030915	LBH	limb bud and heart development homolog (mouse)	0.000572052	1.8056
8034217	NM_000121	EPOR	erythropoietin receptor	0.00089038	1.80302
8039884	NM_020535	KIR2DL5A	killer cell immunoglobulin-like receptor, two domains,	0.0138353	1.80093
7929750	NM_020354	ENTPD7	ectonucleoside triphosphate diphosphohydrolase 7	0.0158149	1.79151
7982377	NM_013372	GREM1	gremlin 1, cysteine knot superfamily, homolog (Xenopus lae	0.00580836	1.78746
7921099	NM_001878	CRABP2	cellular retinoic acid binding protein 2	0.0359018	1.78663
8167815	NM_014599	MAGED2	melanoma antigen family D, 2	0.00496292	1.78208
8034108	NM_024029	YIPF2	Yip1 domain family, member 2	0.000753058	1.77698
7940479	NM_016499	TMEM216	transmembrane protein 216	0.00966689	1.77526
7968711	NR_026745	COG6	component of oligomeric golgi complex 6	0.0136498	1.77489
8010897	NM_001004431	METRNL	meteorin, glial cell differentiation regulator-like	0.0301889	1.77331
7930208	NM_032727	INA	internexin neuronal intermediate filament protein, alpha	0.00154104	1.77151
8129562	NM_001901	CTGF	connective tissue growth factor	0.00788969	1.76947
8161558	NR_002817	AQP7P1	aquaporin 7 pseudogene 1	0.0442446	1.76905
8048926	NM_138402	SP140L	SP140 nuclear body protein-like	0.0143985	1.76873
8122265	NM_006290	TNFAIP3	tumor necrosis factor, alpha-induced protein 3	0.0146818	1.75607

8021470	NM_021127	PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1	0.0055053	1.75421
8038949	NM_001136499	ZNF841	zinc finger protein 841	0.0125107	1.75327
8103951	NM_001995	ACSL1	acyl-CoA synthetase long-chain family member 1	0.00706812	1.75292
8016390	NM_016429	COPZ2	coatomer protein complex, subunit zeta 2	0.00751349	1.75221
8130176	NM_024518	ULBP3	UL16 binding protein 3	0.0435273	1.74743
7894204	---			0.0199515	1.74173
8157270	NM_001859	SLC31A1	solute carrier family 31 (copper transporters), member 1	0.0362952	1.73988
8175023	NM_016032	ZDHHC9	zinc finger, DHHC-type containing 9	0.0100979	1.73808
7973629	NM_001048205	REC8	REC8 homolog (yeast)	0.0136612	1.73763
8169186	NM_017752	TBC1D8B	TBC1 domain family, member 8B (with GRAM domain)	0.00195151	1.73548
7899627	NM_022164	TINAGL1	tubulointerstitial nephritis antigen-like 1	0.0156082	1.73451
7998927	---			0.0207259	1.73047
7972713	NM_004093	EFNB2	ephrin-B2	0.0183569	1.72992
8037301	NM_014400	LYPD3	LY6	0.000706136	1.72794
7895844	---			0.0240262	1.72744
8177788	NM_005516	HLA-E	major histocompatibility complex, class I, E	0.0214438	1.72538
8179103	NM_005516	HLA-E	major histocompatibility complex, class I, E	0.0214438	1.72538
7895136	---			0.0179274	1.72504
8021301	NM_004163	RAB27B	RAB27B, member RAS oncogene family	0.0125159	1.7231
8044417	NM_032824	TMEM87B	transmembrane protein 87B	0.0461803	1.72126
8075310	NM_002309	LIF	leukemia inhibitory factor (cholinergic differentiation fact	0.0073838	1.72105
8107750	NM_130809	PRRC1	proline-rich coiled-coil 1	0.0133017	1.71685
8163257	NM_057159	LPAR1	lysophosphatidic acid receptor 1	0.00377084	1.71676
7930413	NM_004419	DUSP5	dual specificity phosphatase 5	0.00805998	1.71357
8002778	NM_152649	MLKL	mixed lineage kinase domain-like	0.0102754	1.71115
7964460	NM_004083	DDIT3	DNA-damage-inducible transcript 3	0.00854241	1.71061
8117890	NM_005516	HLA-E	major histocompatibility complex, class I, E	0.0248342	1.70697
7906863	NM_003115	UAP1	UDP-N-acteylglucosamine pyrophosphorylase 1	0.00193558	1.70386
8004867	NM_001025579	NDEL1	nudE nuclear distribution gene E homolog (A. nidulans)-	0.00168661	1.70257
7949904	NM_030930	UNC93B	unc-93 homolog B1 (C. elegans)	0.0204066	1.70049

		1			
8097461	NM_012118	CCRN4L	CCR4 carbon catabolite repression 4-like (S. cerevisiae)	0.00390576	1.69944
7997662	NM_014732	KIAA0513	KIAA0513	0.00165634	1.69941
8074931	NM_021916	ZNF70	zinc finger protein 70	0.00517722	1.69924
8133155	NM_003596	TPST1	tyrosylprotein sulfotransferase 1	0.00836217	1.69913
8168875	NM_016607	ARMCX3	armadillo repeat containing, X-linked 3	0.00815738	1.69824
7924450	NM_007207	DUSP10	dual specificity phosphatase 10	0.00311541	1.69682
7995362	NM_133443	GPT2	glutamic pyruvate transaminase (alanine aminotransferase) 2	0.0200675	1.69676
7930074	NM_002502	NFKB2	nuclear factor of kappa light polypeptide gene enhancer in	0.0359613	1.69399
8019316	NM_006907	PYCR1	pyrroline-5-carboxylate reductase 1	0.00107213	1.69071
7902771	NM_016009	SH3GLB1	SH3-domain GRB2-like endophilin B1	0.000467162	1.68797
8089743	NM_212543	B4GALT4	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, poly	0.00736803	1.68716
8000263	NM_153603	COG7	component of oligomeric golgi complex 7	0.00605564	1.68528
8076515	NM_014570	ARFGAP3	ADP-ribosylation factor GTPase activating protein 3	0.000151186	1.68235
7964535	NM_000785	CYP27B1	cytochrome P450, family 27, subfamily B, polypeptide 1	0.00510058	1.68146
8037374	NM_002659	PLAUR	plasminogen activator, urokinase receptor	0.000306382	1.6759
7934920	NM_001127605	LIPA	lipase A, lysosomal acid, cholesterol esterase	0.0172131	1.67441
8107673	NM_023927	GRAMD3	GRAM domain containing 3	0.0129393	1.67109
8061073	---			0.00088551	1.671
7955606	NM_021934	C12orf44	chromosome 12 open reading frame 44	0.0429176	1.67017
7984779	NM_033240	PML	promyelocytic leukemia	0.0380472	1.66718
7936419	NM_018017	C10orf118	chromosome 10 open reading frame 118	0.0471839	1.66625
8174322	NM_024657	MORC4	MORC family CW-type zinc finger 4	0.00816231	1.66378
8129937	NM_006079	CITED2	Cbp	0.0082104	1.66353
8150698	NM_003068	SNAI2	snail homolog 2 (Drosophila)	0.00300937	1.6622
8170635	NM_001080485	ZNF275	zinc finger protein 275	3.37E-05	1.65974
7905754	NM_020452	ATP8B2	ATPase, class I, type 8B, member 2	0.000195066	1.6585
8101723	NM_153757	NAP1L5	nucleosome assembly protein 1-like 5	0.00626464	1.65123
81480	NM_0025	NOV	nephroblastoma overexpressed	0.042079	1.65065

49	14		gene		
7904050	NM_020963	MOV10	Mov10, Moloney leukemia virus 10, homolog (mouse)	0.0117567	1.65061
8082003	NM_018456	EAF2	ELL associated factor 2	0.0131707	1.64886
7958202	NM_018413	CHST11	carbohydrate (chondroitin 4) sulfotransferase 11	0.0355299	1.64455
8174228	NM_153333	TCEAL8	transcription elongation factor A (SII)-like 8	0.00115663	1.64392
8003611	NM_024792	FAM57A	family with sequence similarity 57, member A	0.00257499	1.64362
7986509	ENST00000423248	C15orf51	dynamin 1 pseudogene	0.0148254	1.64291
7986512	ENST00000426730	LOC100289668	similar to dynamin 1	0.0148254	1.64291
7986527	AK125787	C15orf51	dynamin 1 pseudogene	0.0148254	1.64291
8035177	NM_024881	SLC35E1	solute carrier family 35, member E1	0.00664766	1.64227
8028332	NM_004823	KCNK6	potassium channel, subfamily K, member 6	0.00217948	1.6403
8005449	NM_000422	KRT17	keratin 17	0.0404879	1.63956
7936871	NM_000274	OAT	ornithine aminotransferase (gyrate atrophy)	0.0207091	1.6371
7965681	NM_153687	IKIP	IKK interacting protein	0.00203733	1.63581
8079993	NM_007022	CYB561D2	cytochrome b-561 domain containing 2	0.00330491	1.63361
7999594	---			0.0130072	1.62865
7999750	---			0.0130072	1.62865
7993451	---			0.0130072	1.62864
7993586	---			0.0130072	1.62864
7949948	NM_022338	C11orf24	chromosome 11 open reading frame 24	0.00313254	1.62855
8070557	NM_001098402	ZNF295	zinc finger protein 295	0.00222195	1.62733
7994675	NM_181718	ASPHD1	aspartate beta-hydroxylase domain containing 1	0.0375688	1.62674
8119088	NM_078467	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	0.00104408	1.6249
7899173	NM_024887	DHDDS	dehydrodolichyl diphosphate synthase	0.00402619	1.62371
8001317	NM_153029	N4BP1	NEDD4 binding protein 1	0.0112025	1.62164
7938390	NM_001124	ADM	adrenomedullin	0.00256029	1.62016
8168868	NM_016608	ARMCX1	armadillo repeat containing, X-linked 1	0.00931011	1.61847
8123678	NM_183373	C6orf145	chromosome 6 open reading frame 145	0.0123011	1.61688
8000028	NM_173475	DCUN1	DCN1, defective in cullin neddylation 1, domain containi	0.00310913	1.61649

		D3			
7973530	NM_004563	PCK2	phosphoenolpyruvate carboxykinase 2 (mitochondrial)	0.0172048	1.61506
7903719	NM_004037	AMPD2	adenosine monophosphate deaminase 2 (isoform L)	8.86E-05	1.61189
7894865	---			0.0186044	1.60951
8016646	NM_000088	COL1A1	collagen, type I, alpha 1	0.0117365	1.60728
8132218	NM_198428	BBS9	Bardet-Biedl syndrome 9	0.00117111	1.60692
8058450	NM_005279	GPR1	G protein-coupled receptor 1	0.0067493	1.60664
8037387	---			0.0445865	1.60642
8117476	NM_006994	BTN3A3	butyrophilin, subfamily 3, member A3	0.0439699	1.60558
7976812	NR_003232	SNORD113-4	small nucleolar RNA, C	0.0289488	1.60482
8058765	NM_212482	FN1	fibronectin 1	0.0207664	1.60469
8179926	NM_005510	DOM3Z	dom-3 homolog Z (C. elegans)	0.00295378	1.60356
8172280	NM_032591	SLC9A7	solute carrier family 9 (sodium	0.00653811	1.60209
7953936	NM_153022	C12orf59	chromosome 12 open reading frame 59	0.037752	1.60037
8180034	NM_000544	TAP2	transporter 2, ATP-binding cassette, sub-family B (MDR	0.0366037	1.59872
8101675	NM_004827	ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	0.00901537	1.59763
8090433	NM_007283	MGLL	monoglyceride lipase	0.0407016	1.59754
8038919	NM_021632	ZNF350	zinc finger protein 350	0.00869686	1.59437
7995755	NM_018233	OGFOD1	2-oxoglutarate and iron-dependent oxygenase domain contai	0.00526924	1.59271
8025877	NM_022737	LPPR2	lipid phosphate phosphatase-related protein type 2	0.00135577	1.59199
7981046	NM_032036	IFI27L2	interferon, alpha-inducible protein 27-like 2	0.037289	1.59002
8105229	NM_015946	PELO	pelota homolog (Drosophila)	0.00337606	1.58813
8026564	NM_016270	KLF2	Kruppel-like factor 2 (lung)	0.0114714	1.58772
7956551	NM_182947	GEFT	RhoA	0.00582382	1.58681
8127549	NM_012434	SLC17A5	solute carrier family 17 (anion	0.00666331	1.58646
8149705	---			0.00590143	1.5848
8079966	NM_004636	SEMA3B	sema domain, immunoglobulin domain (Ig), short basic doma	0.010011	1.58421
8072678	NM_002133	HMOX1	heme oxygenase (decycling) 1	0.0456024	1.58398
7895534	---			0.0390898	1.58373
7986517	ENST00000423248	C15orf51	dynamin 1 pseudogene	0.014306	1.57898

7986522	ENST00000423248	C15orf51	dynamin 1 pseudogene	0.014306	1.57898
8116780	NM_004415	DSP	desmoplakin	0.016225	1.57722
8030128	NM_014330	PPP1R15A	protein phosphatase 1, regulatory (inhibitor) subunit 1	0.0160235	1.5754
8059413	NM_014689	DOCK10	dedicator of cytokinesis 10	0.00794067	1.57411
8064014	NM_022082	SLC17A9	solute carrier family 17, member 9	0.0242341	1.57337
8178841	NM_018833	TAP2	transporter 2, ATP-binding cassette, sub-family B (MDR	0.0498075	1.57215
8057045	NM_181342	FKBP7	FK506 binding protein 7	0.00175355	1.57177
7894900	---			0.0248689	1.57079
8046680	NM_019091	PLEKHA3	pleckstrin homology domain containing, family A (phospho	0.0346562	1.57078
8045664	NM_177964	LYPD6B	LY6	0.0439725	1.57047
8173340	NM_145119	PJA1	praja ring finger 1	0.00830704	1.57024
7984743	NM_001024736	CD276	CD276 molecule	0.0016776	1.56766
7892945	---			0.0497516	1.56695
7903294	NM_033055	HIAT1	hippocampus abundant transcript 1	0.0185598	1.56655
7995350	---			0.00458105	1.5662
8169249	NM_012216	MID2	midline 2	0.032484	1.5643
8116418	NM_005110	GFPT2	glutamine-fructose-6-phosphate transaminase 2	0.0135359	1.56256
8074644	XM_001725793	FAM108A5	family with sequence similarity 108, member A5	0.0133911	1.5621
7915444	NM_022356	LEPRE1	leucine proline-enriched proteoglycan (leprecan) 1	0.0493732	1.56179
8115851	NM_003714	STC2	stanniocalcin 2	0.014156	1.56151
7903586	NM_020141	TMEM167B	transmembrane protein 167B	0.000932492	1.56044
8039905	NM_020141	TMEM167B	transmembrane protein 167B	0.000932492	1.56044
8164343	NM_001035254	FAM102A	family with sequence similarity 102, member A	0.00276221	1.55965
8143367	NM_207113	SLC37A3	solute carrier family 37 (glycerol-3-phosphate transport	0.0363307	1.55888
8029321	NM_181845	ZNF283	zinc finger protein 283	0.00135351	1.55824
8163116	NM_019114	EPB41L4B	erythrocyte membrane protein band 4.1 like 4B	0.00785756	1.55787
8135172	NM_001126340	ORAI2	ORAI calcium release-activated calcium modulator 2	0.00670067	1.55476

7986515	ENST00000426730	LOC100289668	similar to dynamin 1	0.0365197	1.55344
7986525	ENST00000426730	LOC100289668	similar to dynamin 1	0.0365197	1.55344
8155442	NR_002817	AQP7P1	aquaporin 7 pseudogene 1	0.0409831	1.55267
7950284	NM_002869	RAB6A	RAB6A, member RAS oncogene family	0.00304998	1.54854
7945349	NM_016526	BET1L	blocked early in transport 1 homolog (S. cerevisiae)-like	0.0045749	1.54759
8050474	NM_020905	RDH14	retinol dehydrogenase 14 (all-trans	0.0191569	1.54699
7966600	NM_024959	SLC24A6	solute carrier family 24 (sodium	0.0227445	1.5464
8006779	NM_020876	ARHGAP23	Rho GTPase activating protein 23	0.0487559	1.54539
7915861	NM_145279	MOBKL2C	MOB1, Mps One Binder kinase activator-like 2C (yeast)	0.000536104	1.54495
7921228	NM_001145312	ETV3	ets variant 3	0.0124593	1.54407
8023401	NM_025214	CCDC68	coiled-coil domain containing 68	0.00147721	1.54406
8075182	NM_005080	XBP1	X-box binding protein 1	0.00204883	1.54391
8161488	NR_002817	AQP7P1	aquaporin 7 pseudogene 1	0.0377753	1.54351
8137404	NM_019015	CHPF2	chondroitin polymerizing factor 2	0.0016088	1.54195
7932512	NM_022365	DNAJC1	DnaJ (Hsp40) homolog, subfamily C, member 1	0.00677179	1.54022
8002854	NM_014567	BCAR1	breast cancer anti-estrogen resistance 1	0.0277224	1.53988
8066214	NM_004613	TGM2	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-	0.0328282	1.53961
7915516	NM_201542	MED8	mediator complex subunit 8	0.00189083	1.53929
7986520	AK302717	C15orf51	dynamin 1 pseudogene	0.0178774	1.53856
8052762	NM_002056	GFPT1	glutamine-fructose-6-phosphate transaminase 1	0.049022	1.53706
8075695	NR_027833	APOL3	apolipoprotein L, 3	0.0206615	1.53696
8012896	NM_000304	PMP22	peripheral myelin protein 22	0.047924	1.53297
7904812	NM_006099	PIAS3	protein inhibitor of activated STAT, 3	0.039685	1.53134
7914603	NM_153341	RNF19B	ring finger protein 19B	0.0206321	1.52974
8070255	---			0.0100849	1.52894
8013042	NM_000422	KRT17	keratin 17	0.0296235	1.52784
8049044	NM_025139	ARMC9	armadillo repeat containing 9	0.0214592	1.5265
8075462	NM_080430	SELM	selenoprotein M	0.0408618	1.52425
8082408	NM_013336	SEC61A	Sec61 alpha 1 subunit (S. cerevisiae)	0.00796279	1.52348

		1			
8165866	NM_000351	STS	steroid sulfatase (microsomal), isozyme S	0.0209743	1.52174
8125483	NM_000544	TAP2	transporter 2, ATP-binding cassette, sub-family B (MDR	0.0435644	1.52158
8055261	NM_032144	RAB6C	RAB6C, member RAS oncogene family	0.017297	1.52155
7938012	NM_001003819	TRIM6-TRIM34	TRIM6-TRIM34 readthrough transcript	0.0396147	1.52086
7987192	NM_001042496	SLC12A6	solute carrier family 12 (potassium	0.0371326	1.51558
8086372	NM_017886	ULK4	unc-51-like kinase 4 (C. elegans)	0.0295493	1.51463
7973902	NM_003136	SRP54	signal recognition particle 54kDa	0.0299541	1.51314
8142096	NM_020725	ATXN7L1	ataxin 7-like 1	0.0233305	1.51308
8053158	NM_006302	MOGS	mannosyl-oligosaccharide glucosidase	0.0119387	1.51031
8022506	---			0.000449494	1.51016
8114583	NM_001035235	SRA1	steroid receptor RNA activator 1	0.0291646	1.5098
8063536	NM_003222	TFAP2C	transcription factor AP-2 gamma (activating enhancer bind	0.0371448	1.5088
8010061	NM_182965	SPHK1	sphingosine kinase 1	0.019796	1.50868
8064042	NM_175609	ARFGAP1	ADP-ribosylation factor GTPase activating protein 1	0.00141085	1.50657
8036835	NM_024877	CNTD2	cyclin N-terminal domain containing 2	0.00914046	1.50624
8099850	NM_024943	TMEM156	transmembrane protein 156	0.0235367	1.5043
8053046	NM_003584	DUSP11	dual specificity phosphatase 11 (RNA	0.00136647	1.50419
8084880	NM_005524	HES1	hairy and enhancer of split 1, (Drosophila)	0.00122383	1.50376
7902043	NM_014787	DNAJC6	DnaJ (Hsp40) homolog, subfamily C, member 6	0.00129197	1.4998
8122457	NM_003764	STX11	syntaxin 11	0.000508172	1.49741
8089759	NM_018266	TMEM39A	transmembrane protein 39A	0.0465465	1.49709
8069574	NM_001100420	C21orf91	chromosome 21 open reading frame 91	0.0217217	1.49506
8138602	NM_004403	DFNA5	deafness, autosomal dominant 5	0.0197859	1.49502
8127051	NM_012288	TRAM2	translocation associated membrane protein 2	0.0123828	1.49487
7934122	NM_001142648	SAR1A	SAR1 homolog A (S. cerevisiae)	7.16E-05	1.49419
8115783	NM_005990	STK10	serine	0.0117389	1.49378
7941274	NM_020680	SCYL1	SCY1-like 1 (S. cerevisiae)	0.00171916	1.49237
7978718	NM_006364	SEC23A	Sec23 homolog A (S. cerevisiae)	0.0057885	1.49154
78986	NM_0017	CDA	cytidine deaminase	0.025618	1.48875

55	85				
8077441	NM_003670	BHLHE40	basic helix-loop-helix family, member e40	0.00111911	1.48868
8032909	NM_005817	PLIN3	perilipin 3	0.00095052	1.48842
7926677	NM_001145373	OTUD1	OTU domain containing 1	0.0118065	1.48763
7948399	NM_152716	PATL1	protein associated with topoisomerase II homolog 1 (yeast)	0.00832429	1.48707
7969677	NM_144778	MBNL2	muscleblind-like 2 (Drosophila)	0.0080019	1.48651
7962579	NM_001143668	AMIGO2	adhesion molecule with Ig-like domain 2	0.0481398	1.48649
8118613	NM_006979	SLC39A7	solute carrier family 39 (zinc transporter), member 7	0.0232249	1.48631
8178225	NM_006979	SLC39A7	solute carrier family 39 (zinc transporter), member 7	0.0232249	1.48631
8179525	NM_006979	SLC39A7	solute carrier family 39 (zinc transporter), member 7	0.0232249	1.48631
8038962	NM_001102657	ZNF836	zinc finger protein 836	0.0209872	1.48313
8091458	NM_014445	SERP1	stress-associated endoplasmic reticulum protein 1	0.000470728	1.48197
7895516	---			0.049653	1.48137
7936856	NM_015892	CHST15	carbohydrate (N-acetylgalactosamine 4-sulfate 6-O) sulfot	0.0137447	1.47826
8036333	NM_020951	ZNF529	zinc finger protein 529	0.0347309	1.47753
8147206	NM_003821	RIPK2	receptor-interacting serine-threonine kinase 2	0.0068648	1.47739
8000236	NM_001802	CDR2	cerebellar degeneration-related protein 2, 62kDa	0.0365413	1.47721
8123936	NM_001142393	NEDD9	neural precursor cell expressed, developmentally down-r	0.0315558	1.47715
8111136	NM_001034850	FAM134B	family with sequence similarity 134, member B	0.0240455	1.47449
7933228	NM_001002265	MARCH8	membrane-associated ring finger (C3HC4) 8	0.0289257	1.47421
8166714	NM_198511	LANCL3	LanC lantibiotic synthetase component C-like 3 (bacterial	0.00237135	1.47412
7985767	NM_022767	AEN	apoptosis enhancing nuclease	0.0100363	1.47128
8113623	NM_021649	TICAM2	toll-like receptor adaptor molecule 2	0.0309885	1.46947
8007427	NM_003734	AOC3	amine oxidase, copper containing 3 (vascular adhesion prote	0.0161948	1.46905
7940565	NM_004265	FADS2	fatty acid desaturase 2	0.0431633	1.4688
8051814	NM_006887	ZFP36L2	zinc finger protein 36, C3H type-like 2	0.0213289	1.46839
8166442	NM_014888	FAM3C	family with sequence similarity 3, member C	0.00909144	1.46826
79398	NM_0011		solute carrier family 39 (zinc	0.0116269	1.46723

05	28225	SLC39A 13	transporter), member		
8132031	NM_175887	PRR15	proline rich 15	0.00337133	1.46606
8130556	NM_001024465	SOD2	superoxide dismutase 2, mitochondrial	0.0107964	1.46603
7946860	NM_000352	ABCC8	ATP-binding cassette, sub-family C (CFTR	0.0242074	1.46584
7980316	NM_003239	TGFB3	transforming growth factor, beta 3	0.015186	1.46571
7993453	NM_173803	MPV17L	MPV17 mitochondrial membrane protein-like	0.0180915	1.46426
8007904	NM_004287	GOSR2	golgi SNAP receptor complex member 2	0.00443777	1.46425
8157264	NM_001860	SLC31A 2	solute carrier family 31 (copper transporters), member 2	0.00886202	1.46323
8016708	NM_018509	LRRC59	leucine rich repeat containing 59	0.000976533	1.46242
7907655	---			0.0353914	1.4617
8017378	NM_001915	CYB561	cytochrome b-561	0.0132065	1.46023
8115997	NM_001031677	RAB24	RAB24, member RAS oncogene family	0.00441483	1.46018
8132118	NM_198098	AQP1	aquaporin 1 (Colton blood group)	0.00730437	1.46006
8017675	NR_024386	PLEKHM 1P	pleckstrin homology domain containing, family M (with R	0.00268055	1.45967
8137680	NM_006869	ADAP1	ArfGAP with dual PH domains 1	0.00374341	1.45835
7990839	NM_181900	STARD5	StAR-related lipid transfer (START) domain containing 5	0.0103882	1.45824
8051413	NM_015475	FAM98A	family with sequence similarity 98, member A	0.000615583	1.45679
7979505	NM_005982	SIX1	SIX homeobox 1	0.00299458	1.45552
7907213	NM_152281	GORAB	golgin, RAB6-interacting	0.00534475	1.45434
8097252	NM_152618	BBS12	Bardet-Biedl syndrome 12	0.0143654	1.4515
7974090	NM_005930	CTAGE5	CTAGE family, member 5	0.000873686	1.45108
7961829	NM_005504	BCAT1	branched chain aminotransferase 1, cytosolic	0.0363464	1.45071
7959500	NM_003959	HIP1R	huntingtin interacting protein 1 related	0.00266642	1.45031
8120927	NM_033411	RWDD2A	RWD domain containing 2A	0.00737215	1.4491
7993148	NM_000303	PMM2	phosphomannomutase 2	0.00241959	1.44832
8134452	NM_177455	BHLHA15	basic helix-loop-helix family, member a15	0.00736895	1.4464
8142540	NM_014888	FAM3C	family with sequence similarity 3, member C	0.00950196	1.44554
7965789	NM_001177	ARL1	ADP-ribosylation factor-like 1	0.0346624	1.44513
8024485	NM_015675	GADD45B	growth arrest and DNA-damage-inducible, beta	0.00110486	1.44287
7990657	NM_144572	TBC1D2	TBC1 domain family, member 2B	0.00851092	1.44267

		B			
8171624	NM_001079858	GPR64	G protein-coupled receptor 64	0.0152112	1.44184
8092765	NM_178496	C3orf59	chromosome 3 open reading frame 59	0.000167028	1.44161
8132070	NM_002047	GARS	glycyl-tRNA synthetase	0.0497947	1.44086
7991587	NM_203472	SELS	selenoprotein S	0.0262662	1.43985
7966035	NM_006825	CKAP4	cytoskeleton-associated protein 4	0.000308145	1.43956
8113286	NM_018343	RIOK2	RIO kinase 2 (yeast)	0.00198689	1.43938
7899615	NM_178865	SERINC2	serine incorporator 2	0.0281902	1.4385
8086330	NM_033027	CSRNP1	cysteine-serine-rich nuclear protein 1	0.0290573	1.43836
8049574	NM_080678	UBE2F	ubiquitin-conjugating enzyme E2F (putative)	0.0073903	1.43798
8142102	NM_020725	ATXN7L1	ataxin 7-like 1	0.0416191	1.43738
8127767	NM_022726	ELOVL4	elongation of very long chain fatty acids (FEN1	0.0210205	1.43664
7946201	NM_012402	ARFIP2	ADP-ribosylation factor interacting protein 2	9.55E-05	1.43642
7946142	NM_145040	PRKCDBP	protein kinase C, delta binding protein	0.0490339	1.43638
8128111	NM_016021	UBE2J1	ubiquitin-conjugating enzyme E2, J1 (UBC6 homolog, yeast)	0.00230163	1.43624
8065089	NM_024704	KIF16B	kinesin family member 16B	0.00621045	1.43611
8030339	NM_001459	FLT3LG	fms-related tyrosine kinase 3 ligand	0.0201489	1.43584
7907171	NM_003666	BLZF1	basic leucine zipper nuclear factor 1	0.0191312	1.43549
7904480	---			0.0335502	1.43546
7904963	---			0.0335502	1.43546
7973306	NM_022060	ABHD4	abhydrolase domain containing 4	0.00104865	1.43515
8030899	NM_001010851	ZNF766	zinc finger protein 766	0.0366636	1.43443
8071927	NM_005265	GGT1	gamma-glutamyltransferase 1	0.00683361	1.4341
8047217	NM_025147	COQ10B	coenzyme Q10 homolog B (S. cerevisiae)	0.0297971	1.43398
8097459	---			0.0427072	1.43364
7977216	NM_015656	KIF26A	kinesin family member 26A	0.0172236	1.43247
7930577	NM_033338	CASP7	caspase 7, apoptosis-related cysteine peptidase	0.00321632	1.43237
7983527	NM_153618	SEMA6D	sema domain, transmembrane domain (TM), and cytoplasmic d	0.00928872	1.43061
7894381	---			0.0240954	1.42989
7896317	---			0.00578388	1.42931
7942957	NM_007173	PRSS23	protease, serine, 23	0.0260665	1.42871

7989037	NM_004748	CCPG1	cell cycle progression 1	0.0135205	1.42823
8029306	AK129550	LOC390940	similar to R28379_1	0.0140287	1.42599
8081256	NM_018309	TBC1D23	TBC1 domain family, member 23	0.0380777	1.42389
7893255	---			0.0431449	1.42348
8151281	NM_014294	TRAM1	translocation associated membrane protein 1	0.0382215	1.42278
7963009	NM_144593	RHEBL1	Ras homolog enriched in brain like 1	0.0459415	1.4224
8008113	NM_005831	CALCOCO2	calcium binding and coiled-coil domain 2	0.0262065	1.42182
8138147	NM_030930	UNC93B1	unc-93 homolog B1 (C. elegans)	0.0489471	1.4213
7895566	---			0.0458037	1.42086
8125220	NM_005510	DOM3Z	dom-3 homolog Z (C. elegans)	0.0161411	1.41999
8178699	NM_005510	DOM3Z	dom-3 homolog Z (C. elegans)	0.0161411	1.41999
8158250	NM_016174	CERCAM	cerebral endothelial cell adhesion molecule	0.00030241	1.41976
7915827	NM_006369	LRRC41	leucine rich repeat containing 41	0.0371145	1.41926
8010243	NM_004710	SYNGR2	synaptogyrin 2	0.0031809	1.41859
7956856	NM_001031679	MSRB3	methionine sulfoxide reductase B3	0.0471678	1.41854
8101376	NM_014933	SEC31A	SEC31 homolog A (S. cerevisiae)	0.0163321	1.41796
8013068	NM_178836	PLD6	phospholipase D family, member 6	0.0223686	1.41776
8137091	NM_170686	ZNF398	zinc finger protein 398	0.0221567	1.41717
8011291	---			0.0382053	1.41595
7900438	NM_023070	ZNF643	zinc finger protein 643	0.00608871	1.41586
8077688	NM_153461	IL17RC	interleukin 17 receptor C	0.0388114	1.41546
8058514	---			0.00762016	1.41538
8152215	NM_005655	KLF10	Kruppel-like factor 10	0.00306897	1.41476
8180396	---			0.0448016	1.41354
8088247	NM_001128615	ARHGEF3	Rho guanine nucleotide exchange factor (GEF) 3	0.0289545	1.41135
7922309	NM_003762	VAMP4	vesicle-associated membrane protein 4	0.00842951	1.41113
7941985	NM_006019	TCIRG1	T-cell, immune regulator 1, ATPase, H+ transporting, lyso	0.0127707	1.41105
8099841	NM_006068	TLR6	toll-like receptor 6	0.0429584	1.4108
7902452	NM_174858	AK5	adenylate kinase 5	0.0148583	1.41064

8145027	NM_022749	FAM160B2	family with sequence similarity 160, member B2	0.0375821	1.40951
7933204	NM_007021	C10orf10	chromosome 10 open reading frame 10	0.013556	1.40786
8099897	NM_003359	UGDH	UDP-glucose dehydrogenase	0.0127487	1.40708
8142524	NM_012338	TSPAN12	tetraspanin 12	0.0056855	1.40702
8034227	NM_001161616	RGL3	ral guanine nucleotide dissociation stimulator-like 3	0.00921306	1.40679
7965573	NM_021229	NTN4	netrin 4	0.00115471	1.40652
8099172	NM_001313	CRMP1	collapsin response mediator protein 1	0.0192718	1.40621
7949577	NM_020470	YIF1A	Yip1 interacting factor homolog A (S. cerevisiae)	3.55E-06	1.40547
7985016	NM_153271	SNX33	sorting nexin 33	0.0264172	1.39871
8000284	NM_015044	GGA2	golgi associated, gamma adaptin ear containing, ARF binding	0.0262957	1.39827
8107353	BC104811	ZRSR1	zinc finger (CCCH type), RNA-binding motif and serine	0.0413251	1.39775
8066513	NM_002999	SDC4	syndecan 4	0.0304537	1.39593
8155096	NM_006368	CREB3	cAMP responsive element binding protein 3	0.0143601	1.39541
8094609	BC040452	FAM114A1	family with sequence similarity 114, member A1	0.00879035	1.39468
7894591	---			0.0291469	1.39441
8115210	NM_006058	TNIP1	TNFAIP3 interacting protein 1	0.0419673	1.39423
8172460	NM_001042498	SLC35A2	solute carrier family 35 (UDP-galactose transporter),	0.046339	1.39408
7962327	NM_052885	SLC2A13	solute carrier family 2 (facilitated glucose transporter	0.00677705	1.39399
8115831	NM_004417	DUSP1	dual specificity phosphatase 1	0.0371596	1.39335
7953229	NM_003845	DYRK4	dual-specificity tyrosine-(Y)-phosphorylation regulated ki	0.0196887	1.39214
8052934	NM_032601	MCEE	methylmalonyl CoA epimerase	0.0298664	1.39213
8141150	NM_133436	ASNS	asparagine synthetase	0.0258703	1.39186
8118863	NM_015245	ANKS1A	ankyrin repeat and sterile alpha motif domain containing	0.038953	1.39179
7996341	NM_020786	PDP2	pyruvate dehyrogenase phosphatase catalytic subunit 2	0.0252461	1.39118
8174134	NM_001006938	TCEAL6	transcription elongation factor A (SII)-like 6	0.0142189	1.39046
8030999	NM_018555	ZNF331	zinc finger protein 331	0.00059384	1.38935
8092750	NM_021032	FGF12	fibroblast growth factor 12	0.0493922	1.38913
8162610	NM_033331	CDC14B	CDC14 cell division cycle 14 homolog B (S. cerevisiae)	0.0171087	1.38911
8113914	NM_133372	FNIP1	folliculin interacting protein 1	0.0491409	1.38795

7909642	NM_024749	VASH2	vasohibin 2	0.0191432	1.38739
7952290	NM_012101	TRIM29	tripartite motif-containing 29	0.0321254	1.38724
8133360	NM_001305	CLDN4	claudin 4	0.0284452	1.38495
8105067	NM_000958	PTGER4	prostaglandin E receptor 4 (subtype EP4)	0.0343093	1.38291
8015016	NM_032865	TNS4	tensin 4	0.0437319	1.38258
8006148	NM_004871	GOSR1	golgi SNAP receptor complex member 1	0.0149608	1.38248
8142270	NM_001037132	NRCAM	neuronal cell adhesion molecule	0.0480102	1.38135
7947681	NM_004308	ARHGAP1	Rho GTPase activating protein 1	0.0200214	1.38093
8030980	NR_003699	ZNF525	zinc finger protein 525	0.0199331	1.38058
7945831	NM_020896	OSBPL5	oxysterol binding protein-like 5	0.00263753	1.38041
8063873	NM_144498	OSBPL2	oxysterol binding protein-like 2	0.0279168	1.37968
7933312	NM_001098845	ANXA8L1	annexin A8-like 1	0.0420588	1.37921
8139656	NM_001001555	GRB10	growth factor receptor-bound protein 10	0.0320858	1.37824
8056005	NM_001105	ACVR1	activin A receptor, type I	0.00983218	1.37784
8126556	NM_023932	DLK2	delta-like 2 homolog (Drosophila)	0.015154	1.37741
7892821	---			0.0139733	1.37713
8156043	NM_058179	PSAT1	phosphoserine aminotransferase 1	0.0459172	1.37702
8075332	NM_031937	TBC1D10A	TBC1 domain family, member 10A	0.000799762	1.37668
8168438	NM_145052	UPRT	uracil phosphoribosyltransferase (FUR1) homolog (S. cerevis	0.0135328	1.37613
7906085	NM_170707	LMNA	lamin A	0.00731511	1.37475
7904433	NM_006623	PHGDH	phosphoglycerate dehydrogenase	0.0249752	1.37405
7915882	NM_014774	KIAA0494	KIAA0494	0.0203662	1.37345
8035445	NM_005354	JUND	jun D proto-oncogene	0.0206254	1.3728
8176865	BC032332	PCMTD2	protein-L-isoaspartate (D-aspartate) O-methyltransferase d	0.0243107	1.37208
8177460	BC032332	PCMTD2	protein-L-isoaspartate (D-aspartate) O-methyltransferase d	0.0243108	1.37208
7915718	NM_007170	TESK2	testis-specific kinase 2	0.0362635	1.3707
8131339	NM_003088	FSCN1	fascin homolog 1, actin-bundling protein (Strongylocentrot	0.0371491	1.37021
8082816	NM_021203	SRPRB	signal recognition particle receptor, B subunit	0.017421	1.36973
8006479	NM_013975	LIG3	ligase III, DNA, ATP-dependent	0.0027693	1.36922
8148304	NM_025195	TRIB1	tribbles homolog 1 (Drosophila)	0.00121539	1.36887

8145854	NM_007198	PROSC	proline synthetase co-transcribed homolog (bacterial)	0.0148609	1.36822
8150002	NM_172366	FBXO16	F-box protein 16	0.0173861	1.36695
8009075	NM_025185	TANC2	tetratricopeptide repeat, ankyrin repeat and coiled-coil c	0.0217837	1.36669
8160981	BC004406	KIAA1539	KIAA1539	0.0369639	1.36642
8026163	NM_004907	IER2	immediate early response 2	0.0405565	1.36635
7995797	NM_175617	MT1E	metallothionein 1E	0.0274231	1.3659
8148149	NM_014943	ZHX2	zinc fingers and homeoboxes 2	0.0285838	1.36585
8165183	NM_014866	SEC16A	SEC16 homolog A (S. cerevisiae)	0.0124957	1.36497
8082478	NM_016128	COPG	coatomer protein complex, subunit gamma	0.014916	1.36474
8143919	NM_001003802	SMARCD3	SWI	0.000600106	1.36281
7944850	NM_032811	TBRG1	transforming growth factor beta regulator 1	0.0467609	1.36277
7923131	NM_001142795	DENND1B	DENN	0.0012724	1.36214
7964548	NM_005371	METTL1	methyltransferase like 1	0.031425	1.36205
8010562	NM_006340	BAIAP2	BAI1-associated protein 2	0.0435754	1.3619
7901054	NM_004073	PLK3	polo-like kinase 3 (Drosophila)	0.00550805	1.36082
8002152	NM_005072	SLC12A4	solute carrier family 12 (potassium	0.00821593	1.36062
8078729	NM_002468	MYD88	myeloid differentiation primary response gene (88)	0.0388924	1.36023
8114030	NM_007054	KIF3A	kinesin family member 3A	0.0146852	1.36004
8001410	NM_001012398	AKTIP	AKT interacting protein	0.0458779	1.36002
8045674	NM_194317	LYPD6	LY6	0.0092306	1.35854
7920082	NM_005060	RORC	RAR-related orphan receptor C	0.00391096	1.35849
7909225	NM_001004023	DYRK3	dual-specificity tyrosine-(Y)-phosphorylation regulated	0.00474205	1.35788
8026007	NM_153358	ZNF791	zinc finger protein 791	0.0494768	1.35673
7944285	NM_001655	ARCN1	archain 1	0.0137896	1.35522
7959298	NM_001080825	TMEM120B	transmembrane protein 120B	0.0425395	1.35419
8035905	NM_032139	ANKRD27	ankyrin repeat domain 27 (VPS9 domain)	0.00806292	1.35353
7932960	---			0.0401381	1.35347
7892898	---			0.00708977	1.3534
7979943	NM_033141	MAP3K9	mitogen-activated protein kinase kinase kinase 9	0.0122681	1.35338

8046815	NM_194250	ZNF804A	zinc finger protein 804A	0.0285706	1.3532
8097256	NM_002006	FGF2	fibroblast growth factor 2 (basic)	0.00442741	1.35294
7958216	NM_015275	KIAA1033	KIAA1033	0.0418477	1.35277
8122013	NM_032438	L3MBTL3	l(3)mbt-like 3 (Drosophila)	0.00323586	1.35224
8029894	NM_015603	CCDC9	coiled-coil domain containing 9	0.0473246	1.35193
8135323	NM_021930	RINT1	RAD50 interactor 1	0.0211919	1.35153
7974425	NM_015589	SAMD4A	sterile alpha motif domain containing 4A	0.00609488	1.35148
8111331	NM_022130	GOLPH3	golgi phosphoprotein 3 (coat-protein)	0.00459593	1.35126
7983321	NM_138423	CASC4	cancer susceptibility candidate 4	0.0377104	1.35091
8169683	NM_006777	ZBTB33	zinc finger and BTB domain containing 33	0.0366697	1.35062
8121916	NM_032784	RSPO3	R-spondin 3 homolog (Xenopus laevis)	0.00256095	1.35022
8113214	NM_002064	GLRX	glutaredoxin (thioltransferase)	0.0261354	1.35012
7917728	NM_001006605	FAM69A	family with sequence similarity 69, member A	0.0191944	1.34851
8162260	---			0.0193427	1.34661
7983365	NM_182985	TRIM69	tripartite motif-containing 69	0.0409129	1.34624
8110841	NM_024830	LPCAT1	lysophosphatidylcholine acyltransferase 1	1.81E-05	1.34606
8009502	NM_000891	KCNJ2	potassium inwardly-rectifying channel, subfamily J, member	0.0237733	1.34556
8166747	NM_138780	SYTL5	synaptotagmin-like 5	0.0225737	1.34468
7930194	NM_017649	CNNM2	cyclin M2	0.00384417	1.34459
7991566	NM_018148	LINS1	lines homolog 1 (Drosophila)	0.0293782	1.34428
7957043	NM_006654	FRS2	fibroblast growth factor receptor substrate 2	0.0280919	1.34419
8173848	NM_144657	HDX	highly divergent homeobox	0.0450139	1.34303
8014794	NM_199247	CACNB1	calcium channel, voltage-dependent, beta 1 subunit	0.0247336	1.34266
8020149	NM_003826	NAPG	N-ethylmaleimide-sensitive factor attachment protein, gamma	0.0228223	1.34171
7993624	NM_016524	SYT17	synaptotagmin XVII	0.0440261	1.34142
8032718	NM_001348	DAPK3	death-associated protein kinase 3	0.0187981	1.33971
8045136	NM_032144	RAB6C	RAB6C, member RAS oncogene family	0.00568305	1.33949
8094144	NM_030930	UNC93B1	unc-93 homolog B1 (C. elegans)	0.0290031	1.33938
7979510	NM_017420	SIX4	SIX homeobox 4	0.048412	1.33837

7894081	---			0.0376108	1.33673
8178489				0.0405924	1.33664
8040753	NM_017727	TMEM214	transmembrane protein 214	0.00258725	1.33601
7988212	NM_025165	ELL3	elongation factor RNA polymerase II-like 3	0.0184779	1.3348
7899350	NM_001143912	FAM76A	family with sequence similarity 76, member A	0.0147666	1.33455
8142143	NM_006348	COG5	component of oligomeric golgi complex 5	0.00874676	1.33446
7910611	NM_002245	KCNK1	potassium channel, subfamily K, member 1	0.0152705	1.33408
8111506	NM_144647	CAPSL	calcyphosine-like	0.00439392	1.33399
8019018	NM_003655	CBX4	chromobox homolog 4 (Pc class homolog, Drosophila)	0.0495441	1.33387
7896581	---			0.0413706	1.33185
8175039	NM_001421	ELF4	E74-like factor 4 (ets domain transcription factor)	0.00687541	1.33089
8091972	NM_001105078	MECOM	MDS1 and EVI1 complex locus	0.00626018	1.33071
8121588	NM_013352	DSE	dermatan sulfate epimerase	0.0182881	1.3306
8058985	NM_022453	RNF25	ring finger protein 25	0.0226952	1.32904
8047078	NM_017694	MFSD6	major facilitator superfamily domain containing 6	0.0117252	1.32896
8082133	NM_006810	PDIA5	protein disulfide isomerase family A, member 5	0.0258692	1.3289
8124901	NM_002117	HLA-C	major histocompatibility complex, class I, C	0.0326488	1.32883
7955729	NM_032889	MFSD5	major facilitator superfamily domain containing 5	0.011624	1.32834
7939590	NM_018389	SLC35C1	solute carrier family 35, member C1	0.0237835	1.32737
7897449	NM_025106	SPSB1	splA	0.0152333	1.32698
8072931	NM_014291	GCAT	glycine C-acetyltransferase (2-amino-3-ketobutyrate coenzym	0.0209718	1.3265
7950197	NM_015242	ARAP1	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1	0.00036364	1.32649
8138735	NM_019102	HOXA5	homeobox A5	0.022888	1.32563
8080781	NM_017771	PXK	PX domain containing serine	0.0026734	1.32523
8000716	NM_012410	SEZ6L2	seizure related 6 homolog (mouse)-like 2	0.00802465	1.32522
7919226	NM_006468	POLR3C	polymerase (RNA) III (DNA directed) polypeptide C (62kD)	0.0212257	1.32449
8100547	---			0.0206538	1.32404
7918857	NM_005725	TSPAN2	tetraspanin 2	0.0300277	1.32304
7915392	NM_024503	HIVEP3	human immunodeficiency virus type I enhancer binding prot	0.00534829	1.32291
8039829	NM_014219	KIR2DL2	killer cell immunoglobulin-like receptor, two domains, l	0.0391197	1.32119
8137582	---			0.00424081	1.32114

8068238	NM_207585	IFNAR2	interferon (alpha, beta and omega) receptor 2	0.00592554	1.32058
8131996	NM_182898	CREB5	cAMP responsive element binding protein 5	0.0299232	1.3198
7990345	NM_003612	SEMA7A	semaphorin 7A, GPI membrane anchor (John Milton Hagen blo	0.0268783	1.31969
7947230	NM_170732	BDNF	brain-derived neurotrophic factor	0.0315608	1.31963
7972217	NM_005842	SPRY2	sprouty homolog 2 (Drosophila)	0.00403705	1.31937
8162117	NM_016548	GOLM1	golgi membrane protein 1	0.00756222	1.31883
8065832	NM_015638	TRPC4AP	transient receptor potential cation channel, subfamily C	0.00403133	1.31823
8138370	NM_020319	ANKMY2	ankyrin repeat and MYND domain containing 2	0.00346775	1.3182
8058927	NM_022152	TMBIM1	transmembrane BAX inhibitor motif containing 1	0.0100805	1.31816
8156761	NM_018946	NANS	N-acetylneuraminic acid synthase	0.0340416	1.31812
8167763	NM_022117	TSPYL2	TSPY-like 2	0.0210706	1.31803
8038078	NM_006801	KDELR1	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein rete	0.0224778	1.31615
8022356	NM_001128626	SPIRE1	spire homolog 1 (Drosophila)	0.00790176	1.31613
8089820	NM_153002	GPR156	G protein-coupled receptor 156	0.00962717	1.31575
7894562	---			0.0430363	1.31525
7945803	NM_001014438	CARS	cysteinyl-tRNA synthetase	0.00504071	1.3145
8031297	NM_014218	KIR2DL1	killer cell immunoglobulin-like receptor, two domains, l	0.0386491	1.31429
7929958	NM_033637	BTRC	beta-transducin repeat containing	0.0103385	1.31427
7901746	---			0.0405851	1.31389
8088550	NM_198859	PRICKLE2	prickle homolog 2 (Drosophila)	0.00619609	1.31327
8018793	NM_001081461	JMJD6	jumonji domain containing 6	0.0442678	1.31312
8073573	NM_001002034	FAM109B	family with sequence similarity 109, member B	0.0244178	1.31299
7924701	NM_022735	ACBD3	acyl-Coenzyme A binding domain containing 3	0.0329363	1.31243
8174092	NM_177949	ARMCX2	armadillo repeat containing, X-linked 2	0.0391549	1.31213
7989915	NM_017858	TIPIN	TIMELESS interacting protein	0.0357343	1.31183
8040927	NM_013392	NRBP1	nuclear receptor binding protein 1	0.00292304	1.31181
7939620	NM_152312	GYLTL1B	glycosyltransferase-like 1B	0.00520872	1.31095
8134180	NM_019004	ANKIB1	ankyrin repeat and IBR domain containing 1	0.0125124	1.31089
8134552	NM_005720	ARPC1B	actin related protein 2	0.00954524	1.31059

8161839	NR_023352	C9orf95	chromosome 9 open reading frame 95	0.0436584	1.31057
7966851	NM_016281	TAOK3	TAO kinase 3	0.0394788	1.31044
8048523	NM_006736	DNAJB2	DnaJ (Hsp40) homolog, subfamily B, member 2	0.0235729	1.3101
7983111	NM_003825	SNAP23	synaptosomal-associated protein, 23kDa	0.0130568	1.30985
8044236	NM_181453	GCC2	GRIP and coiled-coil domain containing 2	0.0273537	1.30936
8024078	NM_005224	ARID3A	AT rich interactive domain 3A (BRIGHT-like)	0.0330808	1.3092
8110930	---			0.0261058	1.30919
7942674	NM_015516	TSKU	tsukushin	0.0138627	1.30834
7975687	NM_001024674	LIN52	lin-52 homolog (C. elegans)	0.0346678	1.30827
7947040	NM_194285	SPTY2D1	SPT2, Suppressor of Ty, domain containing 1 (S. cerevisi	2.96E-05	1.30793
8032749	NR_002602	SNORD37	small nucleolar RNA, C	0.00611129	1.30646
8061428	NM_001247	ENTPD6	ectonucleoside triphosphate diphosphohydrolase 6 (putativ	0.0190837	1.30492
8118395	NR_026717	STK19	serine	0.0229817	1.30473
8178164	NR_026717	STK19	serine	0.0229817	1.30473
7970162	NM_015205	ATP11A	ATPase, class VI, type 11A	0.0433172	1.30456
8074569	NR_003267	GGT3P	gamma-glutamyltransferase 3 pseudogene	0.00203573	1.3044
8038225	NM_020904	PLEKHA4	pleckstrin homology domain containing, family A (phospho	0.0424189	1.30411
8065242	AF161557	HSPC072	hypothetical LOC29075	0.0338474	1.30332
8090630	NM_001017395	TMCC1	transmembrane and coiled-coil domain family 1	0.0365405	1.30327
8060225	NM_203346	HDLBP	high density lipoprotein binding protein	0.00330987	1.30269
7920193	NM_178351	LCE1C	late cornified envelope 1C	0.00230353	1.3021
8117081	NM_153042	AOF1	amine oxidase (flavin containing) domain 1	0.0293727	1.30199
8162404	NM_001393	ECM2	extracellular matrix protein 2, female organ and adipocyte	0.00416161	1.30168
8114829	NM_001024947	YIPF5	Yip1 domain family, member 5	0.0416732	1.30131
7893136	---			0.033156	1.30064
7924069	---			0.0368196	1.30052
8094056	NM_053044	HTRA3	HtrA serine peptidase 3	0.0231424	1.30024
8091048	NR_023350	COPB2	coatomer protein complex, subunit beta 2 (beta prime)	0.0104419	1.3002
8037444	NM_004234	ZNF235	zinc finger protein 235	0.0136678	1.29999
8145889	NM_004095	EIF4EBP1	eukaryotic translation initiation factor 4E binding pro	0.043654	1.29915

7948588	NM_004200	SYT7	synaptotagmin VII	0.01382	1.29906
8066822	NM_018837	SULF2	sulfatase 2	0.00923386	1.2986
8142981	NM_001018111	PODXL	podocalyxin-like	0.0151454	1.29825
7914530	NM_001143888	BSDC1	BSD domain containing 1	0.00274293	1.29813
8007154	NM_021939	FKBP10	FK506 binding protein 10, 65 kDa	0.0469035	1.29765
8152606	NM_021021	SNTB1	syntrophin, beta 1 (dystrophin-associated protein A1, 59kD	0.0141009	1.29743
8115664	BC099705	GABRB2	gamma-aminobutyric acid (GABA) A receptor, beta 2	0.0461201	1.29736
8062119	AF348994	MT1JP	metallothionein 1J (pseudogene)	0.0292408	1.29734
8085300	NR_024272	SEC13	SEC13 homolog (S. cerevisiae)	0.00335008	1.29607
8061447	NM_002862	PYGB	phosphorylase, glycogen; brain	0.0209609	1.29552
7990815	NM_001100880	ST20	suppressor of tumorigenicity 20	0.0468373	1.29423
8105267	NM_002203	ITGA2	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 recepto	0.0497974	1.29409
7990757	NM_004390	CTSH	cathepsin H	0.0130675	1.29358
8094378	NM_018323	PI4K2B	phosphatidylinositol 4-kinase type 2 beta	0.00883418	1.29259
8164515	NM_014908	DOLK	dolichol kinase	0.00280326	1.29257
8009255	NM_138363	CCDC45	coiled-coil domain containing 45	0.0122323	1.29247
7945539	NM_024698	SLC25A22	solute carrier family 25 (mitochondrial carrier: glutam	0.0412939	1.29213
8133459	NM_003388	CLIP2	CAP-GLY domain containing linker protein 2	0.00170341	1.29211
8174313	NM_024657	MORC4	MORC family CW-type zinc finger 4	0.000732913	1.29171
8159318	NM_001145638	GPSM1	G-protein signaling modulator 1 (AGS3-like, C. elegans)	0.0315359	1.29136
8156228	NM_001912	CTSL1	cathepsin L1	0.0303447	1.28995
8137693	NM_001031617	COX19	COX19 cytochrome c oxidase assembly homolog (S. cerevis	0.0166908	1.28989
7993713	NM_153208	IQCK	IQ motif containing K	0.0295401	1.28931
7954985	NM_032256	TMEM117	transmembrane protein 117	0.0234876	1.28904
7979710	NM_016445	PLEK2	pleckstrin 2	0.0125924	1.28883
8120279	NM_018214	LRRC1	leucine rich repeat containing 1	0.00349145	1.28874
8008716	NR_002307	MSX2P1	msh homeobox 2 pseudogene 1	0.00933599	1.28869
7956287	NM_005967	NAB2	NGFI-A binding protein 2 (EGR1 binding protein 2)	0.0106001	1.28838
7987048	NM_017762	MTMR10	myotubularin related protein 10	0.0402456	1.28777
8029360	NM_013361	ZNF223	zinc finger protein 223	0.0112592	1.28722
8074316	NR_003267	GGT3P	gamma-glutamyltransferase 3 pseudogene	0.000315879	1.28716

8086555	NM_020347	LZTFL1	leucine zipper transcription factor-like 1	0.0400679	1.28647
8048995	NM_030926	ITM2C	integral membrane protein 2C	0.0244129	1.286
8108099	NM_021982	SEC24A	SEC24 family, member A (S. cerevisiae)	0.0332775	1.28506
8097829	NM_033393	FHDC1	FH2 domain containing 1	0.0287046	1.28499
8033319	NM_005490	SH2D3A	SH2 domain containing 3A	0.0304765	1.28476
8048340	NM_005444	RQCD1	RCD1 required for cell differentiation1 homolog (S. pombe)	0.0304834	1.28473
8101304	NM_152545	RASGEF1B	RasGEF domain family, member 1B	0.00508244	1.28454
8083901	NM_022763	FNDC3B	fibronectin type III domain containing 3B	0.00595963	1.28424
8167125	NM_004651	USP11	ubiquitin specific peptidase 11	0.0111153	1.28418
8057771	NM_003151	STAT4	signal transducer and activator of transcription 4	0.00374463	1.28386
8154656	NM_022160	DMRTA1	DMRT-like family A1	0.00589477	1.2837
7967810	NM_005895	GOLGA3	golgi autoantigen, golgin subfamily a, 3	0.00727048	1.28359
8009476	NM_002758	MAP2K6	mitogen-activated protein kinase kinase 6	0.00468283	1.28301
8053142	NM_033046	RTKN	rhotekin	0.00453045	1.28249
7892779	---			0.00778693	1.28215
8168205	NM_005938	FOXO4	forkhead box O4	0.0304929	1.28177
8065080	---			0.0308072	1.28139
8018558	NM_004035	ACOX1	acyl-Coenzyme A oxidase 1, palmitoyl	0.0422296	1.27755
8036902	NM_013376	SERTAD1	SERTA domain containing 1	0.0215484	1.2775
8108238	NM_001001419	SMAD5	SMAD family member 5	0.0110862	1.27658
8158666	AK128680	HMCN2	hemicentin 2	0.026947	1.27622
8042905	NM_013247	HTRA2	HtrA serine peptidase 2	0.040243	1.27621
8033813	NM_017703	FBXL12	F-box and leucine-rich repeat protein 12	0.0166791	1.27606
8180349	---			0.0281839	1.27561
8061305	NM_018474	C20orf19	chromosome 20 open reading frame 19	0.0266513	1.27557
7995492	NM_001114	ADCY7	adenylate cyclase 7	0.0482877	1.27552
8112182	NM_152622	MIER3	mesoderm induction early response 1, family member 3	0.0350765	1.27545
8051998	NM_139279	MCFD2	multiple coagulation factor deficiency 2	0.0168049	1.27529
8037315	NM_198850	PHLDB3	pleckstrin homology-like domain, family B, member 3	0.00346149	1.27474
7896608	---			0.0476873	1.27352

8070876	NM_015227	POFUT2	protein O-fucosyltransferase 2	0.0333737	1.27321
7995803	AF348994	MT1JP	metallothionein 1J (pseudogene)	0.0473409	1.27297
8035714	NM_016573	GMIP	GEM interacting protein	0.00657851	1.27271
8165639	BC052297	C9orf169	chromosome 9 open reading frame 169	0.000245913	1.2727
8152491	NM_000127	EXT1	exostoses (multiple) 1	0.00728496	1.27243
7918203	NM_001143989	NBPF4	neuroblastoma breakpoint family, member 4	0.0426601	1.27227
8099797	NM_001085399	RELL1	RELT-like 1	0.00866872	1.27221
8171034	NM_005840	SPRY3	sprouty homolog 3 (Drosophila)	0.023595	1.2719
8176955	NM_005840	SPRY3	sprouty homolog 3 (Drosophila)	0.023595	1.2719
8091550	NM_020776	KIAA1328	KIAA1328	0.0253853	1.27166
7916422	NM_152268	PARS2	prolyl-tRNA synthetase 2, mitochondrial (putative)	0.0496431	1.27166
8131135	NM_018641	CHST12	carbohydrate (chondroitin 4) sulfotransferase 12	0.038386	1.27152
8020847	NM_001390	DTNA	dystrobrevin, alpha	0.0338634	1.2714
8074716	NR_003267	GGT3P	gamma-glutamyltransferase 3 pseudogene	0.000909698	1.27122
8073705	NM_181333	PRR5	proline rich 5 (renal)	0.00100051	1.26953
8108166	NM_024715	TXNDC15	thioredoxin domain containing 15	0.0497888	1.26923
7896952	NM_018188	ATAD3A	ATPase family, AAA domain containing 3A	0.0332893	1.26898
7996012	NM_012106	ARL2BP	ADP-ribosylation factor-like 2 binding protein	0.0433107	1.26875
8164235	NM_001002913	PTRH1	peptidyl-tRNA hydrolase 1 homolog (S. cerevisiae)	0.00861868	1.26849
7935746	NM_001001342	BLOC1S2	biogenesis of lysosomal organelles complex-1, subunit	0.0068102	1.26828
8001048	NM_003041	SLC5A2	solute carrier family 5 (sodium	0.0366529	1.26764
8049187	NM_025202	EFHD1	EF-hand domain family, member D1	0.00270143	1.26742
7956524	NM_024779	PIP4K2C	phosphatidylinositol-5-phosphate 4-kinase, type II, gamm	0.000269629	1.26691
7897295	NM_138697	TAS1R1	taste receptor, type 1, member 1	0.00900729	1.26688
8103998	NM_018359	UFSP2	UFM1-specific peptidase 2	0.0257937	1.26674
8061154	NM_003434	ZNF133	zinc finger protein 133	0.0403603	1.26628
7917875	NM_001993	F3	coagulation factor III (thromboplastin, tissue factor)	0.0018578	1.26626
8157203	NM_001015882	DNAJC25	DnaJ (Hsp40) homolog, subfamily C , member 25	0.0244385	1.26611
8018209	NM_015654	NAT9	N-acetyltransferase 9 (GCN5-related, putative)	0.00141721	1.26558

8114536	NM_198282	TMEM173	transmembrane protein 173	0.0281455	1.26542
7938687	NM_005013	NUCB2	nucleobindin 2	0.0212859	1.26465
8005707	NM_002756	MAP2K3	mitogen-activated protein kinase kinase 3	0.0206886	1.26456
8004360	NM_001002914	KCTD11	potassium channel tetramerisation domain containing 11	0.0212824	1.2643
8031732	NM_173631	ZNF547	zinc finger protein 547	0.0136736	1.26427
8162147	NM_024617	ZCCHC6	zinc finger, CCHC domain containing 6	0.0495895	1.26423
8052834	BC005079	C2orf42	chromosome 2 open reading frame 42	0.0152851	1.26411
7984124	NM_031301	APH1B	anterior pharynx defective 1 homolog B (C. elegans)	0.0224002	1.26407
7892989	---			0.0425064	1.26397
7964203	NM_013449	BAZ2A	bromodomain adjacent to zinc finger domain, 2A	0.00918949	1.26387
8084971	NM_152699	SENP5	SUMO1	0.0433533	1.26368
8136954	NM_001130025	FAM115C	family with sequence similarity 115, member C	0.0273981	1.26353
8074593	NM_015672	RIMBP3	RIMS binding protein 3	0.0232312	1.26348
8126770	NM_016593	CYP39A1	cytochrome P450, family 39, subfamily A, polypeptide 1	0.00802986	1.26338
7920725	NM_005698	SCAMP3	secretory carrier membrane protein 3	0.0415875	1.26333
7977273	NM_152328	ADSSL1	adenylosuccinate synthase like 1	0.0472493	1.26285
8051707	NM_003618	MAP4K3	mitogen-activated protein kinase kinase kinase kinase 3	0.046509	1.26197
7971644	NM_020456	C13orf1	chromosome 13 open reading frame 1	0.0184677	1.26181
7912056	NM_198681	PLEKHG5	pleckstrin homology domain containing, family G (with Rh	0.0260917	1.26136
8099721	NM_015187	KIAA0746	KIAA0746 protein	0.0445441	1.26135
7895562	---			0.0244391	1.2613
7894134	---			0.0427552	1.26051
8106303	NM_016218	POLK	polymerase (DNA directed) kappa	0.0238474	1.25973
8154178	NM_004972	JAK2	Janus kinase 2	0.0119685	1.25958
7913768	NM_021258	IL22RA1	interleukin 22 receptor, alpha 1	0.048977	1.2592
8043945	NM_145686	MAP4K4	mitogen-activated protein kinase kinase kinase kinase 4	0.0422148	1.2591
7990417	NM_005697	SCAMP2	secretory carrier membrane protein 2	0.0229855	1.25902
8137414	NM_016118	NUB1	negative regulator of ubiquitin-like proteins 1	0.0289231	1.25789
8072605	NM_001098535	RFPL3	ret finger protein-like 3	0.026698	1.257

8073422	NM_003216	TEF	thyrotrophic embryonic factor	0.0453069	1.25694
7950578	NM_001128620	PAK1	p21 protein (Cdc42	0.0150048	1.25633
7905382	NM_020832	ZNF687	zinc finger protein 687	0.0303121	1.2562
8126360	NM_004275	MED20	mediator complex subunit 20	0.0160716	1.25589
7923453	NM_006618	KDM5B	lysine (K)-specific demethylase 5B	0.0437545	1.25569
7950471	NM_001005285	OR2AT4	olfactory receptor, family 2, subfamily AT, member 4	0.0443285	1.25562
8072328	NM_012429	SEC14L2	SEC14-like 2 (S. cerevisiae)	0.00781805	1.25554
8028674	NM_003169	SUPT5H	suppressor of Ty 5 homolog (S. cerevisiae)	0.00198564	1.25508
7919751	NM_021960	MCL1	myeloid cell leukemia sequence 1 (BCL2-related)	0.0436938	1.25474
8117869	NM_172016	TRIM39	tripartite motif-containing 39	0.0364357	1.25469
8177770	NM_172016	TRIM39	tripartite motif-containing 39	0.0364357	1.25469
8179088	NM_172016	TRIM39	tripartite motif-containing 39	0.0364357	1.25469
8094599	NM_016531	KLF3	Kruppel-like factor 3 (basic)	0.00775072	1.25448
8014233	NM_001104587	SLFN11	schlafen family member 11	0.0402863	1.25445
8156022	AK129705	LOC642947	hypothetical protein LOC642947	2.35E-05	1.25437
7899023	NM_015627	LDLRAP1	low density lipoprotein receptor adaptor protein 1	0.00732412	1.25151
7924549	BC032332	PCMTD2	protein-L-isoaspartate (D-aspartate) O-methyltransferase d	0.0446867	1.2511
8049963	BC032332	PCMTD2	protein-L-isoaspartate (D-aspartate) O-methyltransferase d	0.044687	1.2511
7984112	NM_016530	RAB8B	RAB8B, member RAS oncogene family	0.024616	1.25084
7905043	NR_027002	LOC388692	hypothetical LOC388692	0.00625656	1.25066
8143221	NM_020632	ATP6V0A4	ATPase, H+ transporting, lysosomal V0 subunit a4	0.00469607	1.2506
8028397	NM_174905	FAM98C	family with sequence similarity 98, member C	0.0137177	1.25041

Supplemental Table 4. Clustering analysis of gene ontology terms for genes regulated by Creb3L1 T in HeLa cells.

Gene Ontology term		Fold Enrichment	P-value
Annotation Cluster 1	**Enrichment Score: 9.25**		
Golgi vesicle transport		6.9	2.9E-14
Secretory pathway		3.7	3.0E-9
Secretion by cell		3.3	1.4E-8
secretion		2.8	8.1E-8
Annotation Cluster 2	**Enrichment Score: 5.2**		
Intracellular transport		2.1	1.1E-6
Cellular localization		1.8	1.1E-5
Establishment of cellular localization		1.8	2.1E-5
Annotation Cluster 3	**Enrichment Score: 4.95**		
Response to virus		5.3	3.7E-8
Response to other organism		2.9	6.0E-5
Multi-organism process		2.2	6.4E-4
Annotation Cluster 4	**Enrichment Score: 4.94**		
Protein transport		2.1	2.9E-6
Establishment of protein localization		2	3.6E-6
Protein localization		1.9	1.8E-5
Macromolecule localization		1.8	8.8E-5
Annotation Cluster 5	**Enrichment Score: 4.24**		
PIRSF001733:rfp transforming protein		5.9	1.2E-5
SPIa/RYanodine receptor SPRY		4.7	1.7E-5
B302, (SPRY)-like		4.6	2.2E-5
Butyrophylin-like		5.2	4.2E-5
Domain:B30.2/SPRY		5.5	6.7E-5
SPRY		4	9.7E-5
SPRY-associated		5.6	1.7E-4
PRY		4.7	5.7E-4
Annotation Cluster 6	**Enrichment Score: 4.2**		
Localization		1.4	8.3E-6
Establishment of localization		1.4	1.5E-4
transport		1.4	1.9E-4
Annotation Cluster 7	**Enrichment Score: 3.29**		
Positive regulation of cellular metabolic process		2.1	2.2E-4
Positive regulation of metabolic process		2	3.1E-4
Positive regulation of transcription		2.2	8.4E-4
Positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process		2.1	1.2E-3
Annotation Cluster 8	**Enrichment Score: 3.27**		
Endoplasmic reticulum part		2	1.3E-4
Nuclear envelope-endoplasmic reticulum network		1.9	9.1E-4
Endoplasmic reticulum membrane		1.9	1.4E-3
Annotation Cluster 9	**Enrichment Score: 3.1**		
Regulation of I-kappaB kinase/NF-kappaB cascade		3.7	2.0E-4
Positive regulation of I-kappaB kinase/NF-kappaB cascade		3.7	3.8E-4
I-kappaB kinase/NF-kappaB cascade		3	8.2E-4

Positive regulation of signal transduction		2.5	6.2E-3
Annotation Cluster 10	**Enrichment Score: 3.03**		
Vesicle coat		5.3	9.2E-5
Coated vesicle membrane		4.9	1.6E-4
Golgi-associated vesicle		4.6	2.8E-4
Membrane coat		4.3	5.0E-4
Coated membrane		4.3	5.0E-4
Cytoplasmic vesicle membrane		2.8	5.7E-3
Cytoplasmic vesicle part		2.7	8.1E-3
Vesicle membrane		2.5	1.2E-2

CURRICULUM VITAE FOR Ph.D. CANDIDATES
The Johns Hopkins University School of Medicine

CAITLIN D. HANLON JULY 1, 2015

EDUCATIONAL HISTORY

Ph.D. expected	2015	Program in Biochemistry, Cellular, and Molecular Biology Johns Hopkins University School of Medicine Mentor:Deborah J. Andrew, PhD
BS	2007	Biochemistry, Cellular, and Molecular Biology Ursinus College

OTHER PROFESSIONAL EXPERIENCE

Research rotation	2008	Lab of Mollie Meffert, PhD, JHMI SOM
Research rotation	2007	Lab of Sinisa Urban, PhD, JHMI SOM
Summer Internship	2006	Biotransformation Department, Wyeth Pharmaceuticals
Summer Fellowship	2005	Lab of James Sidie, PhD, Ursinus College

SCHOLARSHIPS AND FELLOWSHIPS

Ruth L Kirschstein National Individual Research Award 5F31DE022233
The Barry Goldwater Scholarship for Excellence in Education

ACADEMIC HONORS AND AWARDS

2012	3rd Place Poster Prize, BCMB Recruiting
2011	2nd Place Poster Prize; BCMB Retreat
2007	Phi Beta Kappa – Tau chapter
2007	Beta Beta Beta, biology honor society

PUBLICATIONS

Hanlon, C.D. and Andrew, D.J. (2015). Outside – In Signaling: a review of the *Drosophila* GPCR family. *Journal of Cell Science* ***in submission***

Chung, S.Y., **Hanlon, C.D**. and Andrew, D.J. (2014). Building and specializing epithelial tubular organs: the Drosophila salivary gland as a model system for revealing how epithelial organs are specified, form and specialize. *WIREs Developmental Biology*. doi: 10.1002/wdev.140.

Fox, R.M., **Hanlon, C.D**. and Andrew, D.J. (2010). The CrebA/Creb3-like transcription factors are major and direct regulators of secretory capacity. *Journal of Cell Biology* 191:479-492.

POSTERS AND ABSTRACTS

Caitlin D. Hanlon, Deborah J. Andrew. The role of Fd64a in salivary gland migration and muscle organization. The 55th Drosophila Research Conference, March 2014, San Diego CA. Abstract 410B

Caitlin D. Hanlon, Deborah J. Andrew. The role of the Fox transcription factor *fd64a* in embryonic salivary gland migration. The 52nd Drosophila Research Conference, April 2011, San Diego CA. Abstract 1011C.

Caitlin D. Hanlon, Melissa Vining, Deborah J. Andrew. The role of the GPCR Tre1 in salivary gland and germ cell migration. The 50th Drosophila Research Conference, March 2009, Chicago IL. Abstract 957C.

Hanlon, C. and Sidie, J. 2006. Time course of anesthetic action and recovery of alkyl alcohols and alkyl aldehydes. J. Penn. Acad. Sci. 79:110.

SERVICE AND LEADERSHIP

2014-2015	Contributor	ASBMB Today
2009-20012	TA, JHMI SOM	Genetics
2011-2013	TA, JHMI SOM	The Fundamentals of Fluorescent and Confocal Microscopy
2008-2010	Tutor, JHMI SOM	Macromolecular Structure and Analysis, Biochemical and Biophysical Principles, Genetics
2010-2012	Policy Representative	BCMB